

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: November 7, 2005, 09:31:57 ; Search time 231 Seconds  
(without alignments)  
3589.672 Million cell updates/sec  
Title: US-09-603-665-5  
Perfect score: 10807  
Sequence: 1 MTSLAQQLQRLALPOSDASL.....CQKTIQQLVTLGPELQSYF 2144  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 2105692 seqs, 386760381 residues  
Total number of hits satisfying chosen parameters: 2105692  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : A\_Geneseq\_16Dec04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10799	99.9	2144	AAB85029	Protein e
2	10799	99.9	2144	ADQ89800	Antagonis
3	10782	99.8	2144	ADS34902	Human aut
4	10222	94.6	2044	ADS34907	Human aut
5	9853	91.2	2036	ADS34905	Human aut
6	9853	91.2	2036	ADS34903	Human aut
7	9853	91.2	2036	ADS34904	Human aut
8	7756.5	71.8	1569	ABG15232	Novel hum
9	6124.5	56.7	1229	ADG66400	Novel hum
10	5026.5	46.5	1149	ADG08012	Novel pro
11	4506	41.7	897	ADG96220	T cell ac
12	4498	41.6	897	ADG96218	T cell ac
13	4494	41.6	897	ADG96152	T cell ac
14	3694	34.2	734	ADM04615	Human pro
15	2585	23.9	515	Aaw54099	Homo sapi
16	2061	19.1	408	ADS34906	Human aut
17	2051.5	19.0	2096	ABB65242	Drosophil
18	2051.5	19.0	2096	ADG89606	Antagonis
19	1779	16.5	349	ADP24166	PRO polyp
20	1778	16.5	349	ADG92729	Human pro
21	1603	14.8	325	ADG09054	Novel pro
22	1086	10.0	1798	ABJ26330	Aspergill
23	1085	10.0	1814	ABJ25730	Aspergill
24	1056.5	9.8	1818	ABP73779	Candida a
25	956	8.8	1769	ABR52974	Protein s

26	956	8.8	1769	7	ADK62658	Disease t
27	663	6.1	135	5	ABP41280	Human ova
28	379	3.5	77	4	AAm17368	Peptide #
29	379	3.5	77	4	ABb36384	Peptide #
30	379	3.5	77	4	AAm29882	Peptide #
31	379	3.5	77	4	ABb31185	Peptide #
32	379	3.5	77	4	ABb21738	Protein #
33	379	3.5	77	4	AAm69542	Human bon
34	379	3.5	77	4	AAm57146	Human bra
35	379	3.5	77	4	ABG51215	Human liv
36	379	3.5	77	4	AAm05057	Peptide #
37	379	3.5	77	5	ABG39167	Human pep
38	309.5	2.9	2228	7	ABR61599	Human gol
39	309.5	2.9	2230	6	ABU07445	Protein d
40	309.5	2.9	2230	7	ABR61600	Human gol
41	308	2.9	2250	7	ABR61601	Human gol
42	308	2.9	2252	7	ABR61602	Human gol
43	295	2.7	4952	8	ADK15818	Human ABC
44	295	2.7	4958	8	ADK15814	Human ABC
45	295	2.7	5058	8	ADK15793	Human ABC

ALIGNMENTS

RESULT 1  
AAB85029  
ID AAB85029 standard; protein; 2144 AA.  
XX  
AC AAB85029;  
XX  
DT 06-AUG-2001 (first entry)  
XX  
DE Protein encoded by BAP28 cDNA consisting of exons 1 to 45.  
XX  
KW BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 1694  
FT Misc-difference /label= Ser or Asn  
FT Misc-difference 1854  
FT Misc-difference /label= Ala or Val  
FT Misc-difference 1967  
FT Misc-difference /label= Asp or Asn  
FT Misc-difference 2017  
FT Misc-difference /label= Gly or Glu  
XX  
WO200100669-A2.  
XX  
04-JAN-2001.  
XX  
23-JUN-2000; 2000WO-IB001183.  
XX  
25-JUN-1999; 98US-0141323P.  
XX  
18-JAN-2000; 2000US-0176880P.  
XX  
(GEST ) GENSET.  
XX  
PI Barry C, Bougueleret L, Chumakov I, Cohen-Akenine A;  
XX  
WPI; 2001-367032/38.  
XX  
N-PSDB; AAF83909, AAF83910.  
XX  
New BAP28 polynucleotides and polypeptides overexpressed in prostate cancer cells for diagnosing prostate tumors, e.g. by hybridization or polymerase chain reaction assays.  
XX  
Claim 14; Page 297-304; 349pp; English.  
XX  
The invention is directed to BAP28 polypeptides, BAP28 polynucleotide sequences and regulatory region located at the 3' and 5' ends of the

CC BAP28 coding region. The BAP28 polypeptides can be expressed by standard  
CC recombinant methodology. BAP28 polynucleotides and polypeptides have been  
CC found to be over expressed in prostate tumour cells, therefore levels of  
CC BAP28 expression and/or activity may be assayed (e.g. by polymerase chain  
CC reaction (PCR)) to diagnose patient suffering from or susceptible to  
CC prostate cancer. Antibodies specific for the BAP28 polypeptides are  
CC useful as diagnostic reagents. Biallelic markers of the BAP28 gene are  
CC useful in genetic analysis. The present sequence represents a protein  
CC encoded by a first cDNA sequence of the BAP28 gene consisting of the  
CC exons 1 to 45  
XX  
SQ Sequence 2144 AA;

Query Match		99.9%;	Score 10799;	DB 4;	Length 2144;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 2144;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MTSLAQQLQRLALPOS	DASLLSRDEVASLLPDKA	ATDRDTAFATGCTGLELLG	60
DB	1	MTSLAQQLQRLALPOS	DASLLSRDEVASLLPDKA	ATDRDTAFATGCTGLELLG	60
QY	61	SFEQFEAPLFSQAKTL	RSVQTKAVNKOLDENISLFI	LHLSPYFLPKPAQKCLEWLI	120
DB	61	SFEQFEAPLFSQAKTL	RSVQTKAVNKOLDENISLFI	LHLSPYFLPKPAQKCLEWLI	120
QY	121	PHIHLYNQDSLACVLP	YHETRI FVRVIOQLKINN	KHRFWLLPVKQSGVPLAKGT	180
DB	121	PHIHLYNQDSLACVLP	YHETRI FVRVIOQLKINN	KHRFWLLPVKQSGVPLAKGT	180
QY	181	HCYKDLGFMDFTCSL	VTKSVKVPFAEYPGSSAOL	RVLLAFYASTIVSALVAED	240
DB	181	HCYKDLGFMDFTCSL	VTKSVKVPFAEYPGSSAOL	RVLLAFYASTIVSALVAED	240
QY	241	KLFPYIQKLGSLPDY	RAATYMIICQISVKVTMENT	FVNSLASQIIKTLTKIPSLIK	300
DB	241	KLFPYIQKLGSLPDY	RAATYMIICQISVKVTMENT	FVNSLASQIIKTLTKIPSLIK	300
QY	301	LSCLIVLLQKRPESIG	KPPFHLNCVPLDITLHGI	SETYDVSPLLRMLPHLVVSI	360
DB	301	LSCLIVLLQKRPESIG	KPPFHLNCVPLDITLHGI	SETYDVSPLLRMLPHLVVSI	360
QY	361	HVTGETEGMDGQIY	KRHEALITKISLAKNND	HLIASLLFEEYISYSSQES	420
DB	361	HVTGETEGMDGQIY	KRHEALITKISLAKNND	HLIASLLFEEYISYSSQES	420
QY	421	LNQOFLPLRLLESKY	PRTLDDVLEHKEIADLKK	QELFHPQVSLSTSGKYQFL	480
DB	421	LNQOFLPLRLLESKY	PRTLDDVLEHKEIADLKK	QELFHPQVSLSTSGKYQFL	480
QY	481	TSMLSLNHLPLAPVR	ILAMNHLKIMKTSKEGV	DESFIKEAVLARIGDNDIV	540
DB	481	TSMLSLNHLPLAPVR	ILAMNHLKIMKTSKEGV	DESFIKEAVLARIGDNDIV	540
QY	541	FEIFKHFSEVITISN	ILNLFQRAELSKNGEW	YVLKIAADILIKEIISND	600
DB	541	FEIFKHFSEVITISN	ILNLFQRAELSKNGEW	YVLKIAADILIKEIISND	600
QY	601	VCLLPFWINDDTSAE	MKIAIYLSKSGICSLH	PLLRGWEEALENVIKSTP	660
DB	601	VCLLPFWINDDTSAE	MKIAIYLSKSGICSLH	PLLRGWEEALENVIKSTP	660
QY	661	NQKMIELLADNINL	GDPSSMLKMWEDLISV	GEESFNLKQVTFHVILSV	720
DB	661	NQKMIELLADNINL	GDPSSMLKMWEDLISV	GEESFNLKQVTFHVILSV	720
QY	721	THPPFAIRPSLLQK	TKKLESVITAVEIPSE	WHIELMDRGI PVELMAH	780
DB	721	THPPFAIRPSLLQK	TKKLESVITAVEIPSE	WHIELMDRGI PVELMAH	780
QY	781	VAVEDSVFLVFLSK	KFIYALKAPKSPFKG	DIWWNPEQLKEDSRDYL	840
DB	781	VAVEDSVFLVFLSK	KFIYALKAPKSPFKG	DIWWNPEQLKEDSRDYL	840

QY	841	DAVHFRVLMKLP	FIKVHLEDVQLFKFC	SVLWTYSSLSNPLNC	SVKTVLTQALYVGCAM	900
DB	841	DAVHFRVLMKLP	FIKVHLEDVQLFKFC	SVLWTYSSLSNPLNC	SVKTVLTQALYVGCAM	900
QY	901	LSSQKTCQKHLAS	ISSPVVTSLLINLGS	PKVEVRRAAIQCLQAL	SGVASPFYLIIDHLI	960
DB	901	LSSQKTCQKHLAS	ISSPVVTSLLINLGS	PKVEVRRAAIQCLQAL	SGVASPFYLIIDHLI	960
QY	961	SKABEITSDAA	YVTDIATLFEEL	OREKLLKSHQKLS	ETLKNLLSCVYSCPSY	1020
DB	961	SKABEITSDAA	YVTDIATLFEEL	OREKLLKSHQKLS	ETLKNLLSCVYSCPSY	1020
QY	1021	VLQGVNGEMVLS	QLLPMAEQLEKIK	QKPTAVLKDEAMV	LHLTKKNEFSVSLNEDPK	1080
DB	1021	VLQGVNGEMVLS	QLLPMAEQLEKIK	QKPTAVLKDEAMV	LHLTKKNEFSVSLNEDPK	1080
QY	1081	SLDIFIKAVHTT	KELYAGMPTIQTAL	EKITTKPPFAAIS	DEKVQOKLLRMLFDLL	1140
DB	1081	SLDIFIKAVHTT	KELYAGMPTIQTAL	EKITTKPPFAAIS	DEKVQOKLLRMLFDLL	1140
QY	1141	SHCAQTVSSV	FKGISVNAEQVRI	ELEPPDKAKPIGT	VOOKRQKMOOKSOD	1200
DB	1141	SHCAQTVSSV	FKGISVNAEQVRI	ELEPPDKAKPIGT	VOOKRQKMOOKSOD	1200
QY	1201	GSYQWVTLIL	LELQHKKLRSPQ	ITLVPFLNLLSR	CLEPLPOEQGNMEYTK	1260
DB	1201	GSYQWVTLIL	LELQHKKLRSPQ	ITLVPFLNLLSR	CLEPLPOEQGNMEYTK	1260
QY	1261	NICQKSLPDG	GKIPKDIIDEEK	FNVELIVQCI	RSEMPQTHHALLL	1320
DB	1261	NICQKSLPDG	GKIPKDIIDEEK	FNVELIVQCI	RSEMPQTHHALLL	1320
QY	1321	HNIMISITFM	GANVNRDITVS	QVINKTVKMWI	PALIQSDSGDSIE	1380
DB	1321	HNIMISITFM	GANVNRDITVS	QVINKTVKMWI	PALIQSDSGDSIE	1380
QY	1381	SVFVDALPHV	PHRRPLTVOL	VDTLGABKFLW	ILLILFEQYVTKTVL	1440
DB	1381	SVFVDALPHV	PHRRPLTVOL	VDTLGABKFLW	ILLILFEQYVTKTVL	1440
QY	1441	EADTEFWF	SVCCFESVQHQI	QSLMNIQYL	LKPEEKETIPKAVS	1500
DB	1441	EADTEFWF	SVCCFESVQHQI	QSLMNIQYL	LKPEEKETIPKAVS	1500
QY	1501	NVETHTSKQ	LHRFKFLSVF	MSQSLSSNNF	LKKVSGGPEILK	1560
DB	1501	NVETHTSKQ	LHRFKFLSVF	MSQSLSSNNF	LKKVSGGPEILK	1560
QY	1561	VAQSMERNAD	KLTVKFWRAL	LSKAYDLLDK	VNALLPTETFI	1620
DB	1561	VAQSMERNAD	KLTVKFWRAL	LSKAYDLLDK	VNALLPTETFI	1620
QY	1621	DLLNNKQONI	SWKKTIVTRFL	KLVDPDLAI	VORKKGEERQAI	1680
DB	1621	DLLNNKQONI	SWKKTIVTRFL	KLVDPDLAI	VORKKGEERQAI	1680
QY	1681	FGAENPDP	PPVPLXTAVK	LIAAPERKEE	KNVLSALLICIAE	1740
DB	1681	FGAENPDP	PPVPLXTAVK	LIAAPERKEE	KNVLSALLICIAE	1740
QY	1741	LTTMKNTSEL	SVSEVYLLSALA	ALQKVETLPH	FISPYLEGILS	1800
DB	1741	LTTMKNTSEL	SVSEVYLLSALA	ALQKVETLPH	FISPYLEGILS	1800
QY	1801	QANIRLTS	LKTKTLATTPR	VLLPAIKKTY	KQIEKWKHMG	1860
DB	1801	QANIRLTS	LKTKTLATTPR	VLLPAIKKTY	KQIEKWKHMG	1860
QY	1861	TSHQSOLTA	FFLEALDFRA	QHSNDLEEVG	KTENCIICDCLV	1920
DB	1861	TSHQSOLTA	FFLEALDFRA	QHSNDLEEVG	KTENCIICDCLV	1920
QY	1921	DWAKTEDAP	KDRLLTFYNL	ADCIAEKLKGL	FTLFAHGLVK	1980

Db 1921 DWAKTEDADPKRLTLTPYNLADCIAEAKGLFTLFAHLVKPFDATLXQVNIKTDFAFD 1980  
Qy 1981 SENDPEKCCLLQFILNCLYKIFLFTQHFISKERAXALMPLVDLENRLGEEKFQER 2040  
Db 1981 SENDPEKCCLLQFILNCLYKIFLFTQHFISKERAXALMPLVDLENRLGEEKFQER 2040  
Qy 2041 VTKHLIPCTIAQSVAMADSLWKPLNYQILLKTRDSSPKVRFAALITVLALAEKLENYI 2100  
Db 2041 VTKHLIPCTIAQSVAMADSLWKPLNYQILLKTRDSSPKVRFAALITVLALAEKLENYI 2100  
Qy 2101 VLLPESIPPLAEIMEDECEVEHQCKTIQOLETVLGEPLQSVF 2144  
Db 2101 VLLPESIPPLAEIMEDECEVEHQCKTIQOLETVLGEPLQSVF 2144

## RESULT 2

ADQ89800

ID ADQ89800 standard; protein; 2144 AA.

XX AC

XX AC

XX AC

DT 21-OCT-2004 (first entry)

XX DE

XX DE Antagonist of cell cycle progression polypeptide #115.

XX KW

XX KW Cytostatic; cancer; cell division cycle; mitosis; meiosis;

XX KW cell cycle progression.

XX OS

XX OS Homo sapiens.

XX PN

XX PN WO2004063362-A2.

XX PD

XX PD 29-JUL-2004.

XX PF

XX PF 31-DEC-2003; 2003WO-CB005635.

XX PR

XX PR 10-JAN-2003; 2003US-0439123P.

XX PR

XX PR 06-MAY-2003; 2003US-0468402P.

XX PA

XX PA (CYCL-) CYCLACEL LTD.

XX PI

XX PI Glover D, Bell G, Frenz L, Midgley C;

XX XX

XX XX WPI; 2004-544089/52.

XX DR

XX DR N-PSDB; ADQ89799.

XX PT

XX PT New cell cycle progression genes and proteins for modulating cell cycle

XX PT progression in cells, for preventing, treating or diagnosing cell

XX PT proliferative diseases (e.g. cancer) or for identifying modulators of

XX PT mitosis or meiosis.

XX PS

XX PS Claim 2; SEQ ID NO 230; 461pp; English.

XX CC

XX CC The present invention relates to a polynucleotide for preventing,

XX CC treating or diagnosing a disease in an individual. The composition or the

XX CC polypeptide, polynucleotide or RNA precursor, or antibody is useful for

XX CC diagnosing, preventing or treating diseases (e.g. cell proliferative

XX CC diseases such as cancer) in an individual. These may also be used for

XX CC identifying substances capable of binding to or modulating the function

XX CC of the polypeptide, capable of affecting the function of the

XX CC corresponding gene, or capable of inhibiting the cell division cycle or

XX CC cell cycle progression, preferably mitosis and/or meiosis. The present

XX CC sequence represents an antagonist of cell cycle progression protein

XX CC sequence.

XX XX

XX XX Sequence 2144 AA;

XX XX

XX XX Query Match 99.9%; Score 10799; DB 8; Length 2144;

XX XX Best Local Similarity 99.8%; Pred. No. 0;

XX XX Matches 2140; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

XX XX

XX XX 1 MTSLAQQLQRLALPOSASLLSRDEVASLLFDPKEAATIDRTAFAGCTGLEELLGIDP 60

Db 1 MTSLAQQLQRLALPOSASLLSRDEVASLLFDPKEAATIDRTAFAGCTGLEELLGIDP 60  
Qy 61 SPEQFEAPLFSOLAKTLERSVOTKAVNKOLDENISLFIHLSPYFLPKPAQCLEWLIHR 120  
Db 61 SPEQFEAPLFSOLAKTLERSVOTKAVNKOLDENISLFIHLSPYFLPKPAQCLEWLIHR 120  
Qy 121 FPHILYNQDSLACVLPYHETRI FVRVIQLLKINNSKHFWFWLLPVKQSGVPLAKGTLLIT 180  
Db 121 FPHILYNQDSLACVLPYHETRI FVRVIQLLKINNSKHFWFWLLPVKQSGVPLAKGTLLIT 180  
Qy 181 HCYKOLGFMDFICSLVTKSVKVPABYPGSSAQLRVLLAFYASTIVSALVAEDVSNIIA 240  
Db 181 HCYKOLGFMDFICSLVTKSVKVPABYPGSSAQLRVLLAFYASTIVSALVAEDVSNIIA 240  
Qy 241 KLFPYIQKGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASOIKTKLTKIPSLIKDG 300  
Db 241 KLFPYIQKGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASOIKTKLTKIPSLIKDG 300  
Qy 301 LSLCLIVLLQKQKPESLGKPPPHLCNVPDLITILHGISTYDVSPLLRYMLPHLVVSIH 360  
Db 301 LSLCLIVLLQKQKPESLGKPPPHLCNVPDLITILHGISTYDVSPLLRYMLPHLVVSIH 360  
Qy 361 HVTGEETGMDGQIYKRHLEAILTKISLKNNDHLASLLFEEYISYSSQEBMSNKVSL 420  
Db 361 HVTGEETGMDGQIYKRHLEAILTKISLKNNDHLASLLFEEYISYSSQEBMSNKVSL 420  
Qy 421 LNEQFLPLIRLLESKYPRTLDVLEHLEKEIADLKKQELFHQFVSLSTGGGQVFLADSD 480  
Db 421 LNEQFLPLIRLLESKYPRTLDVLEHLEKEIADLKKQELFHQFVSLSTGGGQVFLADSD 480  
Qy 481 TSLMLSLNHP LAPVRILAMNHLKIMTKTSKEGVDSFIKEAVLARLGDDNIDVLSAISA 540  
Db 481 TSLMLSLNHP LAPVRILAMNHLKIMTKTSKEGVDSFIKEAVLARLGDDNIDVLSAISA 540  
Qy 541 FEIFKEHFSSEVTISNLLNLFQRAELSKNGEYEVLTAAADILIKBEILSENDQLSNQV 600  
Db 541 FEIFKEHFSSEVTISNLLNLFQRAELSKNGEYEVLTAAADILIKBEILSENDQLSNQV 600  
Qy 601 VCLLPFVVINDDTESAEMKIAIYLSKSGICSLHPLLRGWEALENVIKSTPGKLI GVA 660  
Db 601 VCLLPFVVINDDTESAEMKIAIYLSKSGICSLHPLLRGWEALENVIKSTPGKLI GVA 660  
Qy 661 NQMIELLADNINLGDPSMLKMVEDLISVGEESFNKQKVT FHVILSVLVSCSSLKE 720  
Db 661 NQMIELLADNINLGDPSMLKMVEDLISVGEESFNKQKVT FHVILSVLVSCSSLKE 720  
Qy 721 THFPPAIRVFSLLQKKIKLESVITAVTIPSEWHIELMDRGI PVELWAHYVEELNSTOR 780  
Db 721 THFPPAIRVFSLLQKKIKLESVITAVTIPSEWHIELMDRGI PVELWAHYVEELNSTOR 780  
Qy 781 VAVEDSVFLVSLKFKFIYALKAPKSPKGDITWNPPEQLKEDSRDYLHLLIGLFEMMLNCA 840  
Db 781 VAVEDSVFLVSLKFKFIYALKAPKSPKGDITWNPPEQLKEDSRDYLHLLIGLFEMMLNCA 840  
Qy 841 DAVHFRVLMKLFIKVHLEDVFLQFKFCVSLVMTYSSLSNPLNCSVKTVLQALYVGCAM 900  
Db 841 DAVHFRVLMKLFIKVHLEDVFLQFKFCVSLVMTYSSLSNPLNCSVKTVLQALYVGCAM 900  
Qy 901 LSSQKQCKHOLASISSPVVTSLLINLGSVPKVRRAAIQCLQALSGVSPYLLIDHLI 960  
Db 901 LSSQKQCKHOLASISSPVVTSLLINLGSVPKVRRAAIQCLQALSGVSPYLLIDHLI 960  
Qy 961 SKABEITSDAAVVIQDLATLFEELQREKKLKHOKLSETLKNLLSCVYSCPSYIAKDLMK 1020  
Db 961 SKABEITSDAAVVIQDLATLFEELQREKKLKHOKLSETLKNLLSCVYSCPSYIAKDLMK 1020  
Qy 1021 VLQVNGEMVLSQLLPMAEQLLEKIQEPTAVLKDEAMVHLTLGKYNEFSVSLNEDPK 1080  
Db 1021 VLQVNGEMVLSQLLPMAEQLLEKIQEPTAVLKDEAMVHLTLGKYNEFSVSLNEDPK 1080  
Qy 1081 SIDIFIKAHVHTTKELYAGMPTIOITALEKITYKPFPAALISDEKVOQKLLMFLDVLNCKN 1140

Db 1081 SLIDIFKAVHTTKELYAGMPTIQTALBKITKPFPAAISDEKVOQKLRMLFDLLVNCKN 1140  
QY 1141 SHCAQTVSVFSGISVNAEQVRIELEPPDKAKPLGTVOQKRQKMOQKKSODLESQVBVG 1200  
Db 1141 SHCAQTVSVFSGISVNAEQVRIELEPPDKAKPLGTVOQKRQKMOQKKSODLESQVBVG 1200  
QY 1201 GSYMQRVTLIIELLOHKKKLRSQILVPTLFNLLSRCLPLPQEQGNMEYTKQLILSCLL 1260  
Db 1201 GSYMQRVTLIIELLOHKKKLRSQILVPTLFNLLSRCLPLPQEQGNMEYTKQLILSCLL 1260  
QY 1261 NICQKLSPDGKIPKDIIDBEKFNVELIVQCIRLSEMQTHHALLILGTWAGIPDKVL 1320  
Db 1261 NICQKLSPDGKIPKDIIDBEKFNVELIVQCIRLSEMQTHHALLILGTWAGIPDKVL 1320  
QY 1321 HNIMISFTFMGANVRMLDDTYSFQVINKVVIIPALIQSDSGDSIEVSRNVEEIVWKII 1380  
Db 1321 HNIMISFTFMGANVRMLDDTYSFQVINKVVIIPALIQSDSGDSIEVSRNVEEIVWKII 1380  
QY 1381 SVFVDALPHVPEHRLPLVLQVDTLGAEKFLWILLILFEQYVTKTVLAAAYGEKDAIL 1440  
Db 1381 SVFVDALPHVPEHRLPLVLQVDTLGAEKFLWILLILFEQYVTKTVLAAAYGEKDAIL 1440  
QY 1441 EADTFWFVSVCCFVSQHQIQSLMILQVLLKLPKEKETIPKAVSFNKSESQEMLOVF 1500  
Db 1441 EADTFWFVSVCCFVSQHQIQSLMILQVLLKLPKEKETIPKAVSFNKSESQEMLOVF 1500  
QY 1501 NVETHTSKQLRHFKFLVSFMSQLSSNNFLKVVESGGPEILKGLERLLTTLVLYISA 1560  
Db 1501 NVETHTSKQLRHFKFLVSFMSQLSSNNFLKVVESGGPEILKGLERLLTTLVLYISA 1560  
QY 1561 VAQSMERNADKLTVMFRALLSKAYDLKVNALLPTETTFIPVIRGLVGNPLPSVRRKAL 1620  
Db 1561 VAQSMERNADKLTVMFRALLSKAYDLKVNALLPTETTFIPVIRGLVGNPLPSVRRKAL 1620  
QY 1621 DLLNNKLQONIISWKKTIIVTRFLKLPDILAIIVORKKEGEEQAINRQALYTLKLLCKN 1680  
Db 1621 DLLNNKLQONIISWKKTIIVTRFLKLPDILAIIVORKKEGEEQAINRQALYTLKLLCKN 1680  
QY 1681 FGAENPDPPVPLXKTAVKLIAPEKKEKNVLSGALICIAEVTSTLEALAIPOPLSIMPSL 1740  
Db 1681 FGAENPDPPVPLXKTAVKLIAPEKKEKNVLSGALICIAEVTSTLEALAIPOPLSIMPSL 1740  
QY 1741 LTTMKNTSELVSEVYLLSALAALQKVETLPHFISPYILEGISQVILHEKITSEMGSSAS 1800  
Db 1741 LTTMKNTSELVSEVYLLSALAALQKVETLPHFISPYILEGISQVILHEKITSEMGSSAS 1800  
QY 1801 QANIRLTSUKTLATTLAPRVLLPAIKTYKQIEKNKKNHMGPFMSILOHIGVMKKEEL 1860  
Db 1801 QANIRLTSUKTLATTLAPRVLLPAIKTYKQIEKNKKNHMGPFMSILOHIGVMKKEEL 1860  
QY 1861 TSHQSOLTAFFLEALDFAQHSENDEEVGKTENCIIICLIVAMVVKLSEVTFRPLFFKLF 1920  
Db 1861 TSHQSOLTAFFLEALDFAQHSENDEEVGKTENCIIICLIVAMVVKLSEVTFRPLFFKLF 1920  
QY 1921 DWAKTEDAPKORLLTFYNIADCTAEKLGFTLTFAGHLVKPFDATLXQVNI SKTDPAFD 1980  
Db 1921 DWAKTEDAPKORLLTFYNIADCTAEKLGFTLTFAGHLVKPFDATLXQVNI SKTDPAFD 1980  
QY 1981 SENDPEKCCLLLOFILNCLYKIFLFTDQHFISKERAXALMMPVLDOENRIGCEEKFOER 2040  
Db 1981 SENDPEKCCLLLOFILNCLYKIFLFTDQHFISKERAXALMMPVLDOENRIGCEEKFOER 2040  
QY 2041 VTKHLIPCIAQFSVAMADDSLWKPLNVIQILLKTRDSSPKVRFPAALITVLALAEKLENYI 2100  
Db 2041 VTKHLIPCIAQFSVAMADDSLWKPLNVIQILLKTRDSSPKVRFPAALITVLALAEKLENYI 2100  
QY 2101 VLLPESIPFLAELMEDCEBVEHOCQKTIQOLETVLGEPLQSYF 2144  
Db 2101 VLLPESIPFLAELMEDCEBVEHOCQKTIQOLETVLGEPLQSYF 2144

ID AD34902 standard; protein; 2144 AA.  
XX AC ADS34902;  
XX AC ADS34902;  
DT 16-DEC-2004 (first entry)  
XX DE Human autoimmune disease-related protein - SEQ ID 116.  
XX DE single nucleotide polymorphism detection; SNP detection;  
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;  
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;  
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;  
KW glomerulonephritis; Grave's disease; Sjogren's disease;  
KW primary systemic vasculitis.  
XX OS Homo sapiens.  
XX FN WO2004083403-A2.  
XX PD 30-SEP-2004.  
XX PF 18-MAR-2004; 2004WO-US008461.  
XX PR 18-MAR-2003; 2003US-0455444P.  
XX PR 25-APR-2003; 2003US-0465241P.  
XX PA (APPL-) APPLERA CORP.  
XX PI Cargill M, Begovich AB, Alexander HC;  
XX DR WPI; 2004-728480/71.  
XX DR N-PSDB; ADS34819.  
XX PT New isolated nucleic acid molecule comprises at least 8 contiguous  
PT nucleotides where one of the nucleotides is a single nucleotide  
PT polymorphism (SNP), useful for diagnosing or treating autoimmune  
PT diseases, e.g. rheumatoid arthritis.  
XX PS Claim 12; SEQ ID NO 116; 123pp; English.  
XX CC The invention comprises amino acid and coding sequences containing  
CC genetic polymorphisms associated with an altered risk of developing an  
CC autoimmune disease (e.g. rheumatoid arthritis). The invention further  
CC comprises a method of identifying an individual that has an altered risk  
CC of developing an autoimmune disease, comprising detecting a single  
CC nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA  
CC and protein sequences of the invention are useful for diagnosing and  
CC treating autoimmune diseases, such as: rheumatoid arthritis, type 1  
CC diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory  
CC bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious  
CC anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease, The  
CC myocarditis, Sjogren's disease, or primary systemic vasculitis. The  
CC present amino acid sequence represents a human autoimmune disease-related  
CC protein of the invention. NOTE: The present sequence is not shown in the  
CC specification, but has been retrieved from the WIPO website.  
XX SQ Sequence 2144 AA;  
Query Match 99.8%; Score 10782; DB 8; Length 2144;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2137; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
QY 1 MTSLAQQLQRLALPQSDASLLSRDEVASILLFQPKAATDRDTAFAGCTGLGELGIDP 60  
Db 1 MTSLAQQLQRLALPQSDASLLSRDEVASILLFQPKAATDRDTAFAGCTGLGELGIDP 60  
QY 61 SPEQFEAPLFSOLAKTLERSVQTKAVNKOLDENISLFLIHLSPYFLLPKPAKCLEWLIHR 120  
Db 61 SPEQFEAPLFSOLAKTLERSVQTKAVNKOLDENISLFLIHLSPYFLLPKPAKCLEWLIHR 120  
QY 121 FTHILYNQDSLTIACVLPYTHETRIFFVRVIOQLKINNSKRWFWLLPVKQSGVPLAKGTLIT 180  
Db 121 FTHILYNQDSLTIACVLPYTHETRIFFVRVIOQLKINNSKRWFWLLPVKQSGVPLAKGTLIT 180

181 HCYKOLGFMDFICSLVTKSVKFAEYPGSSAQLRVLLAFYASTIVSALVAEDVSDNIIA 240  
181 HCYKDLGFMDFICSLVTKSVKFAEYPGSSAQLRVLLAFYASTIVSALVAEDVSDNIIA 240  
241 KLPFYIOKGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTITKIPSLIKDG 300  
241 KLPFYIOKGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTITKIPSLIKDG 300  
301 LSLCLIVLLQKQKPEGLKGFPHLCNVPDLITILHGISYDYVSPILRYMPLHPLVSVIHH 360  
301 LSLCLIVLLQKQKPEGLKGFPHLCNVPDLITILHGISYDYVSPILRYMPLHPLVSVIHH 360  
361 HVTGEETEGMDGQIYKXHLBAIITKISLKNLHLLASLLFEBYISYSSQEEEMDSNKNVSL 420  
361 HVTGEETEGMDGQIYKXHLBAIITKISLKNLHLLASLLFEBYISYSSQEEEMDSNKNVSL 420  
421 LNEQFLPLIRLESKYPRTLDDVLEHLEKEIADLKQELPHQFVSLSTSGKYQFLADSD 480  
421 LNEQFLPLIRLESKYPRTLDDVLEHLEKEIADLKQELPHQFVSLSTSGKYQFLADSD 480  
481 TSLMLSINHLPLAPVRILAMNHLKKIMKTSKEGVDESFIKEAVLARLGDDNIDVLSAISA 540  
481 TSLMLSINHLPLAPVRILAMNHLKKIMKTSKEGVDESFIKEAVLARLGDDNIDVLSAISA 540  
541 FEIFKHFSPSEVTISNLLNLPQAEISKNGEYEVILKIAADILIKBEIISENDQLSNQVV 600  
541 FEIFKHFSPSEVTISNLLNLPQAEISKNGEYEVILKIAADILIKBEIISENDQLSNQVV 600  
601 VCLLPFWINNDTEAEKMIKIALYLSKSGICSLHPLLRGWEELALENVIKSTKPKGLIGVA 660  
601 VCLLPFWINNDTEAEKMIKIALYLSKSGICSLHPLLRGWEELALENVIKSTKPKGLIGVA 660  
661 NQKMIELLADNINLGPSSMLKMWEDLISVGEESFNLLKQKVTFFHVLVSVSCSSSLKE 720  
661 NQKMIELLADNINLGPSSMLKMWEDLISVGEESFNLLKQKVTFFHVLVSVSCSSSLKE 720  
721 THPPFAIRVFLSLQKKIKKLESVITAVEIPSEWHIELMLDRGIPVELWAHYVEELNSTQR 780  
721 THPPFAIRVFLSLQKKIKKLESVITAVEIPSEWHIELMLDRGIPVELWAHYVEELNSTQR 780  
781 VAVEDSVFLVSLKKTIFALKAPKSPKGDIMNPEQLKEDSDYLLHLLIGLFEMMLNGA 840  
781 VAVEDSVFLVSLKKTIFALKAPKSPKGDIMNPEQLKEDSDYLLHLLIGLFEMMLNGA 840  
841 DAHFRVLMKLFIKVHLEDVFLQFKFCSVLWTYGSLSNPLNCSVKTVLQALYVGCAM 900  
841 DAHFRVLMKLFIKVHLEDVFLQFKFCSVLWTYGSLSNPLNCSVKTVLQALYVGCAM 900  
901 LSSQKTQCKHQHSLASISSPVVTSLLINLGPSPKVEVRRAAIOCLQALSGVASPPYLIIDHLI 960  
901 LSSQKTQCKHQHSLASISSPVVTSLLINLGPSPKVEVRRAAIOCLQALSGVASPPYLIIDHLI 960  
961 SKAEBITSDAAVYIQLATLFEELQREKKLKHOKLSETLKNLLSCVYSCPSYIAKDMLK 1020  
961 SKAEBITSDAAVYIQLATLFEELQREKKLKHOKLSETLKNLLSCVYSCPSYIAKDMLK 1020  
1021 VLQGVNGEMVLSOLLPMWQLEKIQEPTAVLKDEAMVHLTLGKYNEFSVSLNEDPK 1080  
1021 VLQGVNGEMVLSOLLPMWQLEKIQEPTAVLKDEAMVHLTLGKYNEFSVSLNEDPK 1080  
1081 SLIDIFIKAVHTTKELYAGMPTIQITALEKTIKPPFAAISDEKVOQKLLRMLFDLLVNCN 1140  
1081 SLIDIFIKAVHTTKELYAGMPTIQITALEKTIKPPFAAISDEKVOQKLLRMLFDLLVNCN 1140  
1141 SHCAQTQVSVFVKGISVNAQVRIELEPPDKAKPLGTVOQKRRQKMOOKSODLESVQEVG 1200  
1141 SHCAQTQVSVFVKGISVNAQVRIELEPPDKAKPLGTVOQKRRQKMOOKSODLESVQEVG 1200  
1201 GSYWQVRLTILELLOHKKLRSQIILVPTLFLNLLSRCLBPLPOEQGNMEYTKQILISCLL 1260  
1201 GSYWQVRLTILELLOHKKLRSQIILVPTLFLNLLSRCLBPLPOEQGNMEYTKQILISCLL 1260

1261 NICQKLSPDGKIPKDIILDEEKENVELIVQICIRLSEMPQTHHALLLGTVAGIFPDKVL 1320  
1261 NICQKLSPDGKIPKDIILDEEKENVELIVQICIRLSEMPQTHHALLLGTVAGIFPDKVL 1320  
1321 HNIMSIPTFMGANVMDLDDTYSQVINKTVKQVIPALIQSDSGDSIEVSRNVEIIVVKII 1380  
1321 HNIMSIPTFMGANVMDLDDTYSQVINKTVKQVIPALIQSDSGDSIEVSRNVEIIVVKII 1380  
1381 SVEFDALPHVPEHRRILPILVQVLDTLGAEKFWILLIILFEQVYVTKVTLAAAVGEKDAIL 1440  
1381 SVEFDALPHVPEHRRILPILVQVLDTLGAEKFWILLIILFEQVYVTKVTLAAAVGEKDAIL 1440  
1441 EADTEFWFVCCBFSVQHOIQSLMNTLOYLKLLPBEKEETIPKAVSFNKSSESEEMLOV 1500  
1441 EADTEFWFVCCBFSVQHOIQSLMNTLOYLKLLPBEKEETIPKAVSFNKSSESEEMLOV 1500  
1501 NVETHTSKOLRHFKFLSVFMSOLLSSNNFLKVVESGGEPEILKGLLEERLLETVLGVISA 1560  
1501 NVETHTSKOLRHFKFLSVFMSOLLSSNNFLKVVESGGEPEILKGLLEERLLETVLGVISA 1560  
1561 VAQSMERNADKLTVEKFRALLSKAYDLDKVNALLPTEFIPVIRGLVGNPLPSVRKAL 1620  
1561 VAQSMERNADKLTVEKFRALLSKAYDLDKVNALLPTEFIPVIRGLVGNPLPSVRKAL 1620  
1621 DLLNNKLQONI SWKKTIVTRFLKLVDPDLAIIVORKKKEGEEQAINQRTALYTLKLLCKN 1680  
1621 DLLNNKLQONI SWKKTIVTRFLKLVDPDLAIIVORKKKEGEEQAINQRTALYTLKLLCKN 1680  
1681 FGAENDDPPVVLXTAVKLIAPERKEKXVLSALLCIAEVTSTLEALAIPOPLSLMPSL 1740  
1681 FGAENDDPPVVLXTAVKLIAPERKEKXVLSALLCIAEVTSTLEALAIPOPLSLMPSL 1740  
1741 LTTMKNTSELVSEVYLLSALAALOKVETLPHFISPYLEGILSOVHLEKITSEMGAS 1800  
1741 LTTMKNTSELVSEVYLLSALAALOKVETLPHFISPYLEGILSOVHLEKITSEMGAS 1800  
1801 QANIRLTSUKKTATTAPRVLLPAIKKTYKQIEKNWKNHMGPFMSILOEHIGXMKKEEL 1860  
1801 QANIRLTSUKKTATTAPRVLLPAIKKTYKQIEKNWKNHMGPFMSILOEHIGXMKKEEL 1860  
1861 TSHQSOLTAFFLEALDFAHQSENDELEVYKTENCIIICLVAMVVKLSEVTPRPLPKLF 1920  
1861 TSHQSOLTAFFLEALDFAHQSENDELEVYKTENCIIICLVAMVVKLSEVTPRPLPKLF 1920  
1921 DWAKTEDAPKRLLTFFYNLADCIAEKLKGLFTLFAHGLVKFPADTLKQVNIISKTDEAFD 1980  
1921 DWAKTEDAPKRLLTFFYNLADCIAEKLKGLFTLFAHGLVKFPADTLKQVNIISKTDEAFD 1980  
1981 SENDPEKCCLLQFILNCLYKIFLFDTHFISKERAXALMPLVDQLENRLGGEKQER 2040  
1981 SENDPEKCCLLQFILNCLYKIFLFDTHFISKERAXALMPLVDQLENRLGGEKQER 2040  
2041 VTKHLIPCTIAQFSVAMADDSLMKPNYOITLLKTRDSSPKVRFAALTIVLALAEKLENYI 2100  
2041 VTKHLIPCTIAQFSVAMADDSLMKPNYOITLLKTRDSSPKVRFAALTIVLALAEKLENYI 2100  
2101 VLLPESIPPLAELMEDECEVEHQCKTIQOQLETVLGEPLQSYF 2144  
2101 VLLPESIPPLAELMEDECEVEHQCKTIQOQLETVLGEPLQSYF 2144

## RESULT 4

ADS34907

ID ADS34907 standard; protein; 2044 AA.

XX ADS34907;

AC ADS34907;

XX 16-DEC-2004 (first entry)

XX Human autoimmune disease-related protein - SEQ ID 121.

XX single nucleotide polymorphism detection; SNP detection;  
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;

KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;  
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;  
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;  
XX primary systemic vasculitis.  
OS Homo sapiens.  
XX WO2004083403-A2.  
XX 30-SEP-2004.  
XX 18-MAR-2004; 2004WO-US008461.  
XX 18-MAR-2003; 2003US-0455444P.  
XX 25-APR-2003; 2003US-0465241P.  
XX (APPL-) APPLERA CORP.  
XX Cargill M, Begovich AB, Alexander HC;  
PI WPI; 2004-728480/71.  
DR N-PSDB; ADS34824.  
XX New isolated nucleic acid molecule comprises at least 8 contiguous  
PT nucleotides where one of the nucleotides is a single nucleotide  
PT polymorphism (SNP), useful for diagnosing or treating autoimmune  
PT diseases, e.g. rheumatoid arthritis.  
XX Claim 12; SEQ ID NO 121; 123pp; English.  
PS The invention comprises amino acid and coding sequences containing  
XX genetic polymorphisms associated with an altered risk of developing an  
CC autoimmune disease (e.g. rheumatoid arthritis). The invention further  
CC comprises a method of identifying an individual that has an altered risk  
CC of developing an autoimmune disease, comprising detecting a single  
CC nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA  
CC and protein sequences of the invention are useful for diagnosing and  
CC treating autoimmune diseases, such as: rheumatoid arthritis, type 1  
CC diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory  
CC bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious  
CC anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,  
CC myocarditis, Sjogren's disease, or primary systemic vasculitis. The  
CC present amino acid sequence represents a human autoimmune disease-related  
CC protein of the invention. NOTE: The present sequence is not shown in the  
CC specification, but has been retrieved from the WIPO website.  
XX Sequence 2044 AA;  
SQ  
Query Match 94.6%; Score 10222; DB 8; Length 2044;  
Best Local Similarity 95.0%; Pred. No. 0;  
Matches 2037; Conservative 1; Mismatches 6; Indels 100; Gaps 2;  
QY 1 MTSIAOQLRALPQSDASLLSDEVASLLFDPEKATIDRTAFPAIGCTGLBEELGIDP 60  
DB 1 MTSIAOQLRALPQSDASLLSDEVASLLFDPEKATIDRTAFPAIGCTGLBEELGIDP 60  
QY 61 SFEQFEAPLFSQAKTLERSVQTKAVNKQLDENISLFLIHLSPYFLKPAQKCLEWLIHR 120  
DB 61 SFEQFEAPLFSQAKTLERSVQTKAVNKQLDENISLFLIHLSPYFLKPAQKCLEWLIHR 120  
QY 121 FHIHLNODSLIACVLPYHETRI FVRVIQLLKINNSKHRFWLLIPVKQSGVPLAKGTLIT 180  
DB 121 FHIHLNODSLIACVLPYHETRI FVRVIQLLKINNSKHRFWLLIPVKQSGVPLAKGTLIT 180  
QY 181 HCYKDLGFMDFICSLVTKSVKVAEYPGSSAQLRLVLLAFYASTIVSALVAEEDVSDNIIA 240  
DB 181 HCYKDLGFMDFICSLVTKSVKVAEYPGSSAQLRLVLLAFYASTIVSALVAEEDVSDNIIA 240  
QY 241 KLPFYIQGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKDG 300  
DB 241 KLPFYIQGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKDG 300  
QY 301 LSCLIVLQRPESLGKPPHLCNVPDILTILHGISETYDVSPLLRYMPLPHLVVSIH 360

DB 301 LSCLIVLQRPESLGKPPHLCNVPDILTILHGISETYDVSPLLRYMPLPHLVVSIH 360  
QY 361 HVTGETEGMDQIYKRHLEAILTKISLKNLHDHLLASLLPEEYISYSQSEMDSNKVS 420  
DB 361 HVTGETEGMDQIYKRHLEAILTKISLKNLHDHLLASLLPEEYISYSQSEMDSNKVS 420  
QY 421 LNEQPLIRLLESKYPRTLDVVLEEHKEADLKKOELFHOFVLSLSTGGKYQFLADSD 480  
DB 421 LNEQPLIRLLESKYPRTLDVVLEEHKEADLKKOELFHOFVLSLSTGGKYQFLADSD 480  
QY 481 TSLMLSLNHLAPVRILAMNHLKIMTKTSKEGVDESFKIYKAVLARIAGDNDIVLSAISA 540  
DB 475 -----EGVDESFKIYKAVLARIAGDNDIVLSAISA 504  
QY 541 FEIFKEHFSSEVTISNLLNLFORAELSKNGEYEVVLKIAADILIKEILSENDQLSNQVV 600  
DB 505 FEIFKEHFSSEVTISNLLNLFORAELSKNGEYEVVLKIAADILIKEILSENDQLSNQVV 564  
QY 601 VCLLPFVIVNDDTESAEKIAIYLSKSGICSLHPLLRGWEAEALENVIKSTPKGLIGVA 660  
DB 565 VCLLPFVIVNDDTESAEKIAIYLSKSGICSLHPLLRGWEAEALENVIKSTPKGLIGVA 624  
QY 661 NOKMIELLADNINILGDPSSMLKMVEDLISVGBEESFNLKQKVTFFHVLVSVCCSLKE 720  
DB 625 NOKMIELLADNINILGDPSSMLKM----- 647  
QY 721 THFFFAIRVFSLLQKKIKKLESVITAVEIPSEWHIELMLDRGIPVELMAHYVEELNSTQR 780  
DB 648 -----BIPSEWHIELMLDRGIPVELMAHYVEELNSTQR 680  
QY 781 VAVEDSVFLVPSLKKFIYALKAPKFPKGDINWNPOLKEDSRDYLHLIGLFEMWLNCA 840  
DB 681 VAVEDSVFLVPSLKKFIYALKAPKFPKGDINWNPOLKEDSRDYLHLIGLFEMWLNCA 740  
QY 841 DAVHFRVLMKLFKVHLEDFVOLFVKFCSVLWTYGGSSLNPLNCNVKTQVLTQALYVGCAM 900  
DB 741 DAVHFRVLMKLFKVHLEDFVOLFVKFCSVLWTYGGSSLNPLNCNVKTQVLTQALYVGCAM 800  
QY 901 LSSQKTQCKHQLASISSPVTSLILNLGSPVKEVRRAAITQCLQALSGVASPPFVLIIDHLI 960  
DB 801 LSSQKTQCKHQLASISSPVTSLILNLGSPVKEVRRAAITQCLQALSGVASPPFVLIIDHLI 860  
QY 961 SKAEBITSDAAVIODLATLFEELQREKKLKHOKLSETLKNLSCVYSPSVIAKDLMK 1020  
DB 861 SKAEBITSDAAVIODLATLFEELQREKKLKHOKLSETLKNLSCVYSPSVIAKDLMK 920  
QY 1021 VLQGVNGENVLSQLLPMAEQLEKIQKPTAVLKDEAMVHLTLTKYNEFSVSLNEDPK 1080  
DB 921 VLQGVNGENVLSQLLPMAEQLEKIQKPTAVLKDEAMVHLTLTKYNEFSVSLNEDPK 980  
QY 1081 SLIDIFIKAVHTTKELYAGMPTTIQITALEKITKPFPAAISDEKVOQKLLRMLFDLLVNCN 1140  
DB 981 SLIDIFIKAVHTTKELYAGMPTTIQITALEKITKPFPAAISDEKVOQKLLRMLFDLLVNCN 1040  
QY 1141 SHCAQTVSSVFGKISVNAEOVRIELEPPDKAPLGTVOQKBRQKMOOKSODLESVQEVG 1200  
DB 1041 SHCAQTVSSVFGKISVNAEOVRIELEPPDKAPLGTVOQKBRQKMOOKSODLESVQEVG 1100  
QY 1201 GSWQORVTLILELLOHKKKLRGPQILVPTILFNLLSRCLBPLPOEQGNMEYTKQLISCLL 1260  
DB 1101 GSWQORVTLILELLOHKKKLRGPQILVPTILFNLLSRCLBPLPOEQGNMEYTKQLISCLL 1160  
QY 1261 NICQKLSPOGKIPKDIIDEEKFNVELIVQCIRLSEMPQTHHALLLGLTGVAGIPDKVL 1320  
DB 1161 NICQKLSPOGKIPKDIIDEEKFNVELIVQCIRLSEMPQTHHALLLGLTGVAGIPDKVL 1220  
QY 1321 HNIMSIPTFMGANVMDLDDTYSFOVINKTVKVIIPALIOSDSDGDSIEVSRNVEEIVVKII 1380  
DB 1221 HNIMSIPTFMGANVMDLDDTYSFOVINKTVKVIIPALIOSDSDGDSIEVSRNVEEIVVKII 1280  
QY 1381 SVFVDALPHVPEHRRLLPIILVQLVDTLGAEBKFWLILLILLPEQYVTKTVLAAAYGEKDAIL 1440



Db 1281 SVFVDALPHVPHRRRLPILVQVDTLGAELFIMILLILLPEQVVTKTVLAACGCKDAIL 1340  
Qy 1441 EADTEFWFSVCCFSSVQHQIQSLMNLQYLLKLPKEKEETIPKAVSFNKSESQEMLQVF 1500  
Db 1341 EADTEFWFSVCCFSSVQHQIQSLMNLQYLLKLPKEKEETIPKAVSFNKSESQEMLQVF 1400  
Qy 1501 NVETHTSKQIRHFKFPLSVFMSQILSSNNFLKVVESGSGPEILKGLERLLETVLGVISA 1560  
Db 1401 NVETHTSKQIRHFKFPLSVFMSQILSSNNFLKVVESGSGPEILKGLERLLETVLGVISA 1460  
Qy 1561 VAQSMERNADKLVKFWALLSKAYDLDKVNALLPTETPIPVIRGLVGNPLPSVRKAL 1620  
Db 1461 VAQSMERNADKLVKFWALLSKAYDLDKVNALLPTETPIPVIRGLVGNPLPSVRKAL 1520  
Qy 1621 DLNNKLQNNISWKKTIVTRFLKLVDPDLAIVQRKKKEGEEQAINRQTALYTLKLLCKN 1680  
Db 1521 DLNNKLQNNISWKKTIVTRFLKLVDPDLAIVQRKKKEGEEQAINRQTALYTLKLLCKN 1580  
Qy 1681 FGAENDDPFPVPLXTAVKLIAPERKEKNVLSGALLCIAEVTSTLEALAIQPLSLMPSL 1740  
Db 1581 FGAENDDPFPVPLXTAVKLIAPERKEKNVLSGALLCIAEVTSTLEALAIQPLSLMPSL 1640  
Qy 1741 LTTMKNTSELVSSEVILLSALAALQKVETLPHFISPYLEGILSQVHLEKITSEMGAS 1800  
Db 1641 LTTMKNTSELVSSEVILLSALAALQKVETLPHFISPYLEGILSQVHLEKITSEMGAS 1700  
Qy 1801 QANIRLTSKKTTLATLAPRVLLPAIKKTYQKIEKNKHMGPFFMSILOEHIGXMKKEEL 1860  
Db 1701 QANIRLTSKKTTLATLAPRVLLPAIKKTYQKIEKNKHMGPFFMSILOEHIGXMKKEEL 1760  
Qy 1861 TSHQSOLTAFFLEALDFAHQSENDELEVGKTENCIDCLVAMVKLSEVTRPLPFKLF 1920  
Db 1761 TSHQSOLTAFFLEALDFAHQSENDELEVGKTENCIDCLVAMVKLSEVTRPLPFKLF 1820  
Qy 1921 DWAKTEDAPKDRLLTFYNLADCTAEKLGFLFAGHLVKPPADTLXQVNIKTDEAFD 1980  
Db 1821 DWAKTEDAPKDRLLTFYNLADCTAEKLGFLFAGHLVKPPADTLXQVNIKTDEAFD 1880  
Qy 1981 SENDPEKCCLLQFILNCLYKIFLFTQHFISKERAXALMPLVDQLENLGGEEKFQER 2040  
Db 1881 SENDPEKCCLLQFILNCLYKIFLFTQHFISKERAXALMPLVDQLENLGGEEKFQER 1940  
Qy 2041 VTKHLIPCTAQSVMADSLWKPLNYQILLKTRDSSPKVRFAALITVLAELAKENYI 2100  
Db 1941 VTKHLIPCTAQSVMADSLWKPLNYQILLKTRDSSPKVRFAALITVLAELAKENYI 2000  
Qy 2101 VLLPESIPFLAELMEDECEVEHOCQKTIQOLETVLGEPLOSYF 2144  
Db 2001 VLLPESIPFLAELMEDECEVEHOCQKTIQOLETVLGEPLOSYF 2044

## RESULT 5

ADS34905

ID ADS34905 standard; protein; 2036 AA.

XX ADS34905;

AC ADS34905;

XX 16-DEC-2004 (first entry)

DT Human autoimmune disease-related protein - SEQ ID 119.

DE single nucleotide polymorphism detection; SNP detection;

XX rheumatoid arthritis; type 1 diabetes; multiple sclerosis;

KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;

KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;

KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;

KW Primary systemic vasculitis.

XX Homo sapiens.

XX WO2004083403-A2.

XX 30-SEP-2004.

PD

XX

PF

XX

PR

XX

PR

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

18-MAR-2004; 2004WO-US008461.

18-MAR-2003; 2003US-0455444P.

25-APR-2003; 2003US-0465241P.

(APPL-) APPLERA CORP.

Cargill M, Begovich AB, Alexander HC;

WPI; 2004-728480/71.

N-PSDB; ADS34822.

New isolated nucleic acid molecule comprises at least 8 contiguous

nucleotides where one of the nucleotides is a single nucleotide

polymorphism (SNP), useful for diagnosing or treating autoimmune

diseases, e.g. rheumatoid arthritis.

Claim 12; SEQ ID NO 119; 123pp; English.

The invention comprises amino acid and coding sequences containing

genetic polymorphisms associated with an altered risk of developing an

autoimmune disease (e.g. rheumatoid arthritis). The invention further

comprises a method of identifying an individual that has an altered risk

of developing an autoimmune disease, comprising detecting a single

nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA

and protein sequences of the invention are useful for diagnosing and

treating autoimmune diseases, such as: rheumatoid arthritis, type 1

diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory

bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious

anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,

myocarditis, Sjogren's disease, or primary systemic vasculitis. The

present amino acid sequence represents a human autoimmune disease-related

protein of the invention. NOTE: The present sequence is not shown in the

specification, but has been retrieved from the WIPO website.

Sequence 2036 AA;

Query Match 91.2%; Score 9853; DB 8; Length 2036;

Best Local Similarity 91.5%; Pred. No. 0;

Matches 1978; Conservative 14; Mismatches 26; Indels 144; Gaps 7;

Qy 1 MTSLAQQLQRLALPQSDASLLSRDEVASLLFPFKAAATIDRTAFAGTGLLELIGIDP 60

Db 1 MTSLAQQLQRLALPQSDASLLSRDEVASLLFPFKAAATIDRTAFAGTGLLELIGIDP 60

Qy 61 SPEQPEAPLFSOLAKTLERSVOTKAVNKOLDENISLFLIHLSPYFLKPAQKCLEWLIHR 120

Db 61 SPEQPEAPLFSOLAKTLERSVOTKAVNKOLDENISLFLIHLSPYFLKPAQKCLEWLIHR 120

Qy 121 FHIHLYNQDSLIAACVLPYHETRIFVRVIOQLKINNSKRWFWLLPVKQSGVPLAKGTLIT 180

Db 121 FHIHLYNQDSLIAACVLPYHETRIFVRVIOQLKINNSKRWFWLLPVKQSGVPLAKGTLIT 180

Qy 181 HCYKDLGFMDFICSIVTKSVKVFABYPGSSAQLRVLLAFYASTIVSALVAEDVSNIIA 240

Db 181 HCYKDLGFMDFICSIVTKSVKVFABYPGSSAQLRVLLAFYASTIVSALVAEDVSNIIA 240

Qy 241 KLFPIYIOKGLKSSLPDYRAATYMIICQISVKVTM---ENTFVNYSLASQ-1IKTLTKIPS- 295

Db 241 KLFPIYIOKGLKSSLPDYRAATYMIICQISVKVTM---ENTFVNYSLASQ-1IKTLTKIPS- 295

Qy 296 -----LIKDGLSCLIVLLQKPKESLKGKPPHLCNVPDLITILHGISETYD 342

Db 296 -----LIKDGLSCLIVLLQKPKESLKGKPPHLCNVPDLITILHGISETYD 342

Qy 343 VSPLRLYMLPHLVSIHHVTGTEGMDGQIYKQHLEAILTKISLKNLDHLLASLLE 402

Db 343 VSPLRLYMLPHLVSIHHVTGTEGMDGQIYKQHLEAILTKISLKNLDHLLASLLE 402

Qy 394 EYISYSSQEMDSNKNVSLNEQFLPLRLLESKYPTLDVLEELKEIADLKKQLFPHQ 462

Db 394 EYISYSSQEMDSNKNVSLNEQFLPLRLLESKYPTLDVLEELKEIADLKKQLFPHQ 462

QY	463	FVSLSTSGGKYQFLADSDTSLMLSLNHLAPVRILAMNHLKKIMTKTSKEGVDESFIKEAV	522
Db	455	FVSLSTSGGKYQ-----EGVDESFIKEAV	478
QY	523	LARLGDNDIVLVAISAFEPFKHEPSSSEVTISNLNLFQRAELSKNGWEYEVLKIADI	582
Db	479	LARLGDNDIVLVAISAFEPFKHEPSSSEVTISNLNLFQRAELSKNGWEYEVLKIADI	538
QY	583	LIKEEILSENDOLSNQVWVCLLPFVIVNDDTBSAEMKIAIYLSKSGICSLHPLLRGWE	642
Db	539	LIKEEILSENDOLSNQVWVCLLPFVIVNDDTBSAEMKIAIYLSKSGICSLHPLLRGWE	598
QY	643	ALENVIKSTKPGKLGIGVANQKMIELLADNINLGDPSMLKMVEDLISVGEESFNLKQKV	702
Db	599	ALENVIKSTKPGKLGIGVANQKMIELLADNINLGDPSMLKM-----	639
QY	703	TFHVILSVLVSCESSLKETHFPFAIRVFSLLQKKIKLSVITAVEIPSEWHIELMLDRG	762
Db	640	-----BIPSEWHIELMLDRG	654
QY	763	IPVELWAHYVEELNSTORVAVEDSVFLVSLKXFIYALKAPKSPFKGDIWMNPEQLKEDS	822
Db	655	IPVELWAHYVEELNSTORVAVEDSVFLVSLKXFIYALKAPKSPFKGDIWMNPEQLKEDS	714
QY	823	RDYLHLILIGLFEMMLNGADAVHRVRLMKLPIKVHLEDVDFQLFKFCSVLTYTGSSLSNPLN	882
Db	715	RDYLHLILIGLFEMMLNGADAVHRVRLMKLPIKVHLEDVDFQLFKFCSVLTYTGSSLSNPLN	774
QY	883	CSVKTVLQDALYVGCAMLSSQKTOCKHOLASISSPWTSSLINLNGSPVKVEVRAAIOCL	942
Db	775	CSVKTVLQDALYVGCAMLSSQKTOCKHOLASISSPWTSSLINLNGSPVKVEVRAAIOCL	834
QY	943	QALSGVASPFYLIIDHLISKAEBITSDAAVYIODLATLFEELQREKKLSHQKLSLTKN	1002
Db	835	QALSGVASPFYLIIDHLISKAEBITSDAAVYIODLATLFEELQREKKLSHQKLSLTKN	894
QY	1003	LLSCVTSYCPSYIAKDLMKVLOGVNGEMVLSQLPMAEOLELKEOKEPTAVLKDEAMVHL	1062
Db	895	LLSCVTSYCPSYIAKDLMKVLOGVNGEMVLSQLPMAEOLELKEOKEPTAVLKDEAMVHL	954
QY	1063	TLGKYNFYSVLINDEPKSLDIFIKAVHTTKELYAGMPTQITALEKTIKPPFAAISDEK	1122
Db	955	TLGKYNFYSVLINDEPKSLDIFIKAVHTTKELYAGMPTQITALEKTIKPPFAAISDEK	1014
QY	1123	VQOKLLRMLFDLLVNCNKSQAOTVSSVFKGISVNAEQVRIELEPPDKAPLGTVOQKRR	1182
Db	1015	VQOKLLRMLFDLLVNCNKSQAOTVSSVFKGISVNAEQVRIELEPPDKAPLGTVOQKRR	1074
QY	1183	QKMOOKSODLESVQEVGGSFWQVTVLILELQHKKLRSPQILVPTFLNLLSRCLEPLP	1242
Db	1075	QKMOOKSODLESVQEVGGSFWQVTVLILELQHKKLRSPQILVPTFLNLLSRCLEPLP	1134
QY	1243	QEQGNMEYTKQILILSCLNITCOKLSPDGGKIPKIDILDEEKFNVELIVQCIRLSEMPQTHH	1302
Db	1135	QEQGNMEYTKQILILSCLNITCOKLSPDGGKIPKIDILDEEKFNVELIVQCIRLSEMPQTHH	1194
QY	1303	HALLILGTVAGIPDKVLHNIMSIPTFWGANVWMLDDTYSFOVINTKTVKMWIPALIQSDS	1362
Db	1195	HALLILGTVAGIPDKVLHNIMSIPTFWGANVWMLDDTYSFOVINTKTVKMWIPALIQSDS	1254
QY	1363	GDSIEVSRNVBEIIVKKIISVFVDALPHVPEHRRILPILVQLVDTLGAEKFLWILLILFEQ	1422
Db	1255	GDSIEVSRNVBEIIVKKIISVFVDALPHVPEHRRILPILVQLVDTLGAEKFLWILLILFEQ	1314
QY	1423	YVTKTVLAAAYGEKDAILEADTEFWFSVCCEFSVQHOIQSLMNILOVLLKLPKEKETIP	1482
Db	1315	YVTKTVLAAAYGEKDAILEADTEFWFSVCCEFSVQHOIQSLMNILOVLLKLPKEKETIP	1374
QY	1483	KAVSFNKSESQEMLOQVFNVEHTHSKOLRHFKFLSVFSMSQLLSSNNFLKXVSVESGGPEI	1542
Db	1375	KAVSFNKSESQEMLOQVFNVEHTHSKOLRHFKFLSVFSMSQLLSSNNFLKXVSVESGGPEI	1434

Qy	1543	LKGLERLL	ETVLGYISAVAQSMERNADKLTKVFWPALLSKAYDLDLDKNALLPTFTTFFP	1602
Db	1435	LKGLERLL	ETVLGYISAVAQSMERNADKLTKVFWPALLSKAYDLDLDKNALLPTFTTFFP	1494
Qy	1603	VIRGLVGN	PLPSVRRKALDLDNNKLOQNTISWKKTIIVTRFLKLPDLDLAIIVORKKKEGEE	1662
Db	1495	VIRGLVGN	PLPSVRRKALDLDNNKLOQNTISWKKTIIVTRFLKLPDLDLAIIVORKKKEGEE	1554
Qy	1663	QAINROTAL	YTLKLLCKNFGAENPDPFVPLXTAVKLIAPERKEEKNVLGSALLCAEVT	1722
Db	1555	QAINROTAL	YTLKLLCKNFGAENPDPFVPLXTAVKLIAPERKEEKNVLGSALLCAEVT	1614
Qy	1723	STLEALAIP	QPLPSLMTSLTMKNTSELVSSEVYLISALAAOKQVETLPHFISPYLEGI	1782
Db	1615	STLEALAIP	QPLPSLMTSLTMKNTSELVSSEVYLISALAAOKQVETLPHFISPYLEGI	1674
Qy	1783	LSQVIHLEKIT	SEMGASQANIIRLTSIKKTLATTLAPRVLLPAIKKTYKQIEKNWKNHMG	1842
Db	1675	LSQVIHLEKIT	SEMGASQANIIRLTSIKKTLATTLAPRVLLPAIKKTYKQIEKNWKNHMG	1734
Qy	1843	PFMSILQEH	IGXMKKEBLTSHQSLTAFFLEALDFRAQSENDELEVGGKTENCIIIDCLVA	1902
Db	1735	PFMSILQEH	IGXMKKEBLTSHQSLTAFFLEALDFRAQSENDELEVGGKTENCIIIDCLVA	1794
Qy	1903	MVVKLSVET	FRPLFFFKLFDWAKTEDAPKORLLTFYNLADCIABKLGFLTFAGHLVKPF	1962
Db	1795	MVVKLSVET	FRPLFFFKLFDWAKTEDAPKORLLTFYNLADCIABKLGFLTFAGHLVKPF	1854
Qy	1963	ADTLXQVNI	SKTDEAPDSENDPEKCLLIQFILNCLYKIFLFDTHFISKEKAXALMMP	2022
Db	1855	ADTLXQVNI	SKTDEAPDSENDPEKCLLIQFILNCLYKIFLFDTHFISKEKAXALMMP	1914
Qy	2023	LVDQENRLG	EEKFQSERVTKHLIPCIQAQSVAMADDSLWKPLNYQILLKTRDSSPKRVF	2082
Db	1915	LVDQENRLG	EEKFQSERVTKHLIPCIQAQSVAMADDSLWKPLNYQILLKTRDSSPKRVF	1974
Qy	2083	AALITVLAL	AEKLEKENVIVILLPSIPFLABELMEDECEVEHQCKTTIQOLETVLGEPLQS	2142
Db	1975	AALITVLAL	AEKLEKENVIVILLPSIPFLABELMEDECEVEHQCKTTIQOLETVLGEPLQS	2034
Qy	2143	YF	2144	
Db	2035	YF	2036	
RESULT 6				
ADS34903				
ID	ADS34903 standard; protein; 2036 AA.			
XX				
AC	ADS34903;			
XX				
DT	16-DEC-2004 (first entry)			
XX				
DE	Human autoimmune disease-related protein - SEQ ID 117.			
XX				
KW	single nucleotide polymorphism detection; SNP detection;			
KW	rheumatoid arthritis; type 1 diabetes; multiple sclerosis;			
KW	systemic lupus erythematosus; inflammatory bowel disease; psoriasis;			
KW	thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;			
KW	glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;			
KW	primary systemic vasculitis.			
OS	Homo sapiens.			
XX				
PN	WO2004083403-A2.			
XX				
PD	30-SEP-2004.			
XX				
PF	18-MAR-2004; 2004WO-US008461.			
XX				
PR	18-MAR-2003; 2003US-045544P.			
PR	25-APR-2003; 2003US-0465241P.			
XX				



PA	(APPL-) APPLERA CORP.
XX	Cargill M, Begowich AB, Alexander HC;
P1	WPI: 2004-728480/71.
DR	N-PADB; ADS34820.
XX	New isolated nucleic acid molecule comprises at least 8 contiguous
PT	nucleotides where one of the nucleotides is a single nucleotide
PT	polymorphism (SNP), useful for diagnosing or treating autoimmune
PT	diseases, e.g. rheumatoid arthritis.
XX	Claim 12; SEQ ID NO 117; 123pp; English.
PS	The invention comprises amino acid and coding sequences containing
XX	genetic polymorphisms associated with an altered risk of developing an
CC	autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC	comprises a method of identifying an individual that has an altered risk
CC	of developing an autoimmune disease, comprising detecting a single
CC	nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC	and protein sequences of the invention are useful for diagnosing and
CC	treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC	diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC	bowel diseases, psoriasis, chondritis, celiac disease, pernicious
CC	anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC	rheumatoid arthritis, Sjogren's disease, or primary systemic vasculitis. The
CC	present amino acid sequence represents a human autoimmune disease-related
CC	protein of the invention. NOTE: The present sequence is not shown in the
CC	specification, but has been retrieved from the WIPO website.
XX	Sequence 2036 AA;
SQ	Query Match           91.2%; Score 9853; DB 8; Length 2036; Best Local Similarity   91.5%; Pred. No. 0; Matches 1978; Conservative 14; Mismatches 26; Indels 144; Gaps 7  OY         1   MTSLAQQLRLAPOSASLASLSRDEVASLLFPDPKEAAITIDRTAFAGICTGLSELLGIDP 60 Db          1   MTSLAQQLRLAPOSASLASLSRDEVASLLFPDPKEAAITIDRTAFAGICTGLSELLGIDP 60  OY         61   SPEOEAPLFSQAKTLERSVQTAVNKOLDENISLFILHSPYFLKPKAQCLEWLHR 120 Db          61   SPEOEAPLFSQAKTLERSVQTAVNKOLDENISLFILHSPYFLKPKAQCLEWLHR 120  OY         121 FHIHLYNODSLIACVLPHETRIFRVVIQLKINNSKHRWFLLPVKQSGVPKLAKTGLTI 180 Db          121 FHIHLYNODSLIACVLPHETRIFRVVIQLKINNSKHRWFLLPVKQSGVPKLAKTGLTI 180  OY         181 HCYKDGFMDPFICSLVTKSVKVFAYEFGSSAQLRVLLAFYASTIVSALVAEDVSDNIIA 240 Db          181 HCYKDGFMDPFICSLVTKSVKVFAYEFGSSAQLRVLLAFYASTIVSALVAEDVSDNIIA 240  OY         241 KLPFYIQKXSSLPDYAATYMIICOISKVTHM---ENTFVNLSAQ--IIKLTWKIPS- 295 Db          241 KLPFYIQKSL-----YLTRTEQTVSLRHQNIDOBLSFDQGCVKLLDSAPAE 289  OY         296 -----LIKDGSLCLLVLLQRQPESIGKKPPHLCNVDPDLITILHGISETYD 342 Db          296 TEAREPWKVCIEIRNGAS-----QRPFHPCNVDPDLITILHGISETYD 334  OY         343 VSPLLRYMLPHLAVSIHHVTGEETEKGMOGOIKYRHLLEAILTKISLKNNLDHLLASLLFE 402 Db          335 VSPILLHYMLPHLAVSIHHVTGEETEKGMOGOIKYRHLLEAILTKISLKNNLDHLLASLLFE 394  OY         403 EYISYSSEEMDNKNVSLLNEOFPLIRLESKYPRTLDVVLEHKETADLKQBLFHQ 462 Db          395 EYISYSSEEMDNKNVSLLNEOFPLIRLESKYPRTLDVVLEHKETADLKQBLFHQ 454  OY         463 FVSLSTSGCKYQFLADSOTSLMSLSLNHPAVRILAMNHCLKIMTKTSKGVDESFTKEAV 522 Db          455 FVSLSTSGCKYQ-----EGVDSESFTKEAV-----EGVDSESFTKEAV 478  OY         523 LARLGDDNDIVVLSAISAFEIFKEHFSESFTVINSLNLFORABLSKNGEWLYELKAADI 582

Db 1495 VIRGLVGNPLPSVRRKALDNLNKKLQONISWKKTVITRFLKVLDPDALLVQRKKGEGBE 1554  
Qy 1663 QAINRQTALYTLKLLCKNFGAENPDPPFVPLXAVKLIAPERKEKNVLSALLCIAEVT 1722  
Db 1555 QAINRQTALYTLKLLCKNFGAENPDPPFVPLXAVKLIAPERKEKNVLSALLCIAEVT 1614  
Qy 1723 STLEALAIQPLSLMPSLLTTWKNTSELVSSEVYLLSALAALQKVETLPHFISPLYEGI 1782  
Db 1615 STLEALAIQPLSLMPSLLTTWKNTSELVSSEVYLLSALAALQKVETLPHFISPLYEGI 1674  
Qy 1783 LSQVHLEKITSEMGSASQANIRLTSLKKTTLATTLAPRVLPAIKKTYKQIEKNWKNHG 1842  
Db 1675 LSQVHLEKITSEMGSASQANIRLTSLKKTTLATTLAPRVLPAIKKTYKQIEKNWKNHG 1734  
Qy 1843 PFMISILQEHIGMKKEELTSHOSQLTAFLEALDFAHQSENDEEVGKTCNCIIDCLVA 1902  
Db 1735 PFMISILQEHIGMKKEELTSHOSQLTAFLEALDFAHQSENDEEVGKTCNCIIDCLVA 1794  
Qy 1903 MVVKLSEVTRFLPFKLPWAKTEDAPKORLLITFYNLADCIABKLGLETLFAGHLVKPF 1962  
Db 1795 MVVKLSEVTRFLPFKLPWAKTEDAPKORLLITFYNLADCIABKLGLETLFAGHLVKPF 1854  
Qy 1963 ADTLXOVNISKTDFAFFDSENDEPEKCLLLOFTLNCLYKIFLFDTOHFTSKERAXALMMP 2022  
Db 1855 ADTLXOVNISKTDFAFFDSENDEPEKCLLLOFTLNCLYKIFLFDTOHFTSKERAXALMMP 1914  
Qy 2023 LVDQLENRLGGEKEFQERVTKHLIPCIAQFSVAMADDSLWKPLNYQILLKTRDSSPKVRF 2082  
Db 1915 LVDQLENRLGGEKEFQERVTKHLIPCIAQFSVAMADDSLWKPLNYQILLKTRDSSPKVRF 1974  
Qy 2083 AALITVLAELKAKENYIVLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLOS 2142  
Db 1975 AALITVLAELKAKENYIVLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLOS 2034  
Qy 2143 YF 2144  
Db 2035 YF 2036

RESULT 7  
ADS34904  
ID ADS34904 standard; protein; 2036 AA.  
XX ADS34904;  
AC ADS34904;  
XX  
DT 16-DEC-2004 (first entry)  
XX Human autoimmune disease-related protein - SEQ ID 118.  
XX  
KW single nucleotide polymorphism detection; SNP detection;  
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;  
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;  
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;  
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;  
KW primary systemic vasculitis.  
XX  
OS Homo sapiens.  
XX  
PN WO2004083403-A2.  
XX  
PD 30-SEP-2004.  
XX  
PF 18-MAR-2004; 2004WO-US008461.  
XX  
PR 18-MAR-2003; 2003US-0455444P.  
PR 25-APR-2003; 2003US-0465241P.  
XX  
XX (APPL-) APPLERA CORP.  
XX  
PI Cargill M, Begovich AB, Alexander HC;  
XX  
XX WPI; 2004-728480/71.  
DR N-PSDB; ADS34821.  
DR

XX New isolated nucleic acid molecule comprises at least 8 contiguous  
PT nucleotides where one of the nucleotides is a single nucleotide  
PT polymorphism (SNP), useful for diagnosing or treating autoimmune  
PT diseases, e.g. rheumatoid arthritis.  
XX  
PS Claim 12; SEQ ID NO 118; 123pp; English.  
XX  
CC The invention comprises amino acid and coding sequences containing  
CC genetic polymorphisms associated with an altered risk of developing an  
CC autoimmune disease (e.g. rheumatoid arthritis). The invention further  
CC comprises a method of identifying an individual that has an altered risk  
CC of developing an autoimmune disease, comprising detecting a single  
CC nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA  
CC and protein sequences of the invention are useful for diagnosing and  
CC treating autoimmune diseases, such as: rheumatoid arthritis, type 1  
CC diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory  
CC bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious  
CC anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,  
CC myocarditis, Sjogren's disease, or primary systemic vasculitis. The  
CC present amino acid sequence represents a human autoimmune disease-related  
CC protein of the invention. NOTE: The present sequence is not shown in the  
CC specification, but has been retrieved from the WIPO website.  
XX  
SQ Sequence 2036 AA;  
Query Match 91.2%; Score 9853; DB 8; Length 2036;  
Best Local Similarity 91.5%; Pred. No. 0;  
Matches 1978; Conservative 14; Mismatches 26; Indels 144; Gaps 7;  
Qy 1 MTSLAQQLQRLALPOSDASLSRDEVASLLPDKAANTIDRTAFAGCTGEEELLGIDP 60  
Db 1 MTSLAQQLQRLALPOSDASLSRDEVASLLPDKAANTIDRTAFAGCTGEEELLGIDP 60  
Qy 61 SPEQFEAPLFSQAKTLERSVOTKAVNKOLDENISLFIHLSPYFLPKPAQCLEWLIHR 120  
Db 61 SPEQFEAPLFSQAKTLERSVOTKAVNKOLDENISLFIHLSPYFLPKPAQCLEWLIHR 120  
Qy 121 FHIHLYNQDSLACVLPYHETRI FVRVIQLKINNSKRWFWLLPVKQSGVPLAKGTLIT 180  
Db 121 FHIHLYNQDSLACVLPYHETRI FVRVIQLKINNSKRWFWLLPVKQSGVPLAKGTLIT 180  
Qy 181 HCYKOLGFMDFICSLVTKSVKFAEYVGSASQRLVLAFAVSTIVSALVAEDVSDNIIA 240  
Db 181 HCYKOLGFMDFICSLVTKSVKFAEYVGSASQRLVLAFAVSTIVSALVAEDVSDNIIA 240  
Qy 241 KLFPIYQKGLKSLPDYRAATYMIICQISVKVTM---ENTFVNSLASQ-IITLTAKIPS- 295  
Db 241 KLFPIYQKSL-----YLTGRRTQTVSLRHQNIQDQSLFDQGWKLLDSAPAE 289  
Qy 296 -----LIKDGLSCLIVLLQROKPSLGKPPPHLCNVDPDLTILHGISETYD 342  
Db 296 TEAREBPWEKVCVTELRNGAS-----QRPPLHLCNVDPDLTILHGISETYD 334  
Qy 343 VSPLLHYMLPHLVVSIHHVTGTEGMDGQIYKRLHLEAILTKISLKNLDDLHLLSLFEE 402  
Db 335 VSPLLHYMLPHLVVSIHHVTGTEGMDGQIYKRLHLEAILTKISLKNLDDLHLLSLFEE 394  
Qy 403 EYISYSSQEEEMDSNKVSLNEQFLPLIRLESKYPRTLDVLEEHLKEIADLKQELFHQ 462  
Db 395 EYISYSSQEEEMDSNKVSLNEQFLPLIRLESKYPRTLDVLEEHLKEIADLKQELFHQ 454  
Qy 463 FVSLSTSGKYQFLADSDTSLMLSNHPLAPVRLAMNHLKIMKTSKEGVDESFIKEAV 522  
Db 455 FVSLSTSGKYQ-----EGVDESFIKEAV 478  
Qy 523 LARLGDNDIDVLSAISAFEIFEKHPSSSVTISNLLNLFORAEKSNKGWYEVYLKAAADI 582  
Db 479 LARLGDNDIDVLSAISAFEIFEKHPSSSVTISNLLNLFORAEKSNKGWYEVYLKAAADI 538  
Qy 583 LIKEITLSENDQLSNOVVVCLLPFVFINNDTSESAPMKIAIYLSKSGICSLHPLLRGWE 642  
Db 539 LIKEITLSENDQLSNOVVVCLLPFVFINNDTSESAPMKIAIYLSKSGICSLHPLLRGWE 598

QY 643 ALENVIKSTKPGKLGIVANGQKMIELLADNINLGDPSMLQWBDLISVGEESFNLKQKV 702  
DB 599 ALENVIKSTKPGKLGIVANGQKMIELLADNINLGDPSMLKM----- 639  
QY 703 TPFVILSVLSCSSSLKETHFPFAIRVFSLLQKKIKKLSBVI TAVBIPSEWHIELMDRG 762  
DB 640 -----BIPSEWHIELMDRG 654  
QY 763 IPVELWAHYVEELNSQRTAVEDSVFLVSLKFKFIYALKAPKSPFKGDIWNPEQLKEDS 822  
DB 655 IPVELWAHYVEELNSQRTAVEDSVFLVSLKFKFIYALKAPKSPFKGDIWNPEQLKEDS 714  
QY 823 RYDLHLILGLFEMMLNGADAVHPRVLMKLFPIKVHLEDVQLFKFCSVLMTYSSLSNPLN 882  
DB 715 RYDLHLILGLFEMMLNGADAVHPRVLMKLFPIKVHLEDVQLFKFCSVLMTYSSLSNPLN 774  
QY 883 CSVKTVLQTOALXVGCAMLSQKQCHOLASTSSPVVTSLLINLGSPPKVEVRRRAIOCL 942  
DB 775 CSVKTVLQTOALXVGCAMLSQKQCHOLASTSSPVVTSLLINLGSPPKVEVRRRAIOCL 834  
QY 943 QALSGVASPFYLIIDHLISKAEBITSDAAYVIOQLATLFEELQREKKLKHQKLSLTKN 1002  
DB 835 QALSGVASPFYLIIDHLISKAEBITSDAAYVIOQLATLFEELQREKKLKHQKLSLTKN 894  
QY 1003 LLSCVTSCPSYIAKOLMKVLQGVNGEMVLSOLLPMABQLLEKIQKPTAVLKDEAMVHL 1062  
DB 895 LLSCVTSCPSYIAKOLMKVLQGVNGEMVLSOLLPMABQLLEKIQKPTAVLKDEAMVHL 954  
QY 1063 TLGKYNFVSLLNEDPKSLDPIKAVHTTKELYAGMPTIOITALEKIKTPPFAAISDEK 1122  
DB 955 TLGKYNFVSLLNEDPKSLDPIKAVHTTKELYAGMPTIOITALEKIKTPPFAAISDEK 1014  
QY 1123 VQOKLRLMFDLLVNCNKHCAQTSSVFSGISVNAEQVRIELEPPDKAKPLGTVOQKR 1182  
DB 1015 VQOKLRLMFDLLVNCNKHCAQTSSVFSGISVNAEQVRIELEPPDKAKPLGTVOQKR 1074  
QY 1183 QKMQOKKSQDLRSVQGVSGYQWVTLILELQHKKKLRSQILVPTLNFLLSRCLEPLP 1242  
DB 1075 QKMQOKKSQDLRSVQGVSGYQWVTLILELQHKKKLRSQILVPTLNFLLSRCLEPLP 1134  
QY 1243 QKMGMEYTKQILSLCLNLCOKLSPDGKIPKIDLEDEKFNVELVQICIRSEMPQTHH 1302  
DB 1135 QKMGMEYTKQILSLCLNLCOKLSPDGKIPKIDLEDEKFNVELVQICIRSEMPQTHH 1194  
QY 1303 HALLLGTGAGIPDPKVLHNIMSIFTFMGANVRLDDTYSFQVINKTVQWVIPALIQSDS 1362  
DB 1195 HALLLGTGAGIPDPKVLHNIMSIFTFMGANVRLDDTYSFQVINKTVQWVIPALIQSDS 1254  
QY 1363 GDSIEVSRNVEEIVVKIISVFDVALPHVPEHRLPILVOLVDTLGAEKFLWILLILLPQ 1422  
DB 1255 GDSIEVSRNVEEIVVKIISVFDVALPHVPEHRLPILVOLVDTLGAEKFLWILLILLPQ 1314  
QY 1423 VYTKTVLAAAYGEKDAILEADTEFPVSCCEFSVQHOIQSLMNILOYLLKLPKEKEETIP 1482  
DB 1315 VYTKTVLAAAYGEKDAILEADTEFPVSCCEFSVQHOIQSLMNILOYLLKLPKEKEETIP 1374  
QY 1483 KAVSFNKSSESQEMLOQVFNVEHTSKQLRHFKFLSVSFMSSQLSSNNFLKVVESGGPEI 1542  
DB 1375 KAVSFNKSSESQEMLOQVFNVEHTSKQLRHFKFLSVSFMSSQLSSNNFLKVVESGGPEI 1434  
QY 1543 LKGLBERLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDLDKVNALLPTETFTIP 1602  
DB 1435 LKGLBERLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDLDKVNALLPTETFTIP 1494  
QY 1603 VIRGLVGNPLPSVRRKALDLNKLQONISWKKTIIVTRFLKLPDILAIVQRKKXGESE 1662  
DB 1495 VIRGLVGNPLPSVRRKALDLNKLQONISWKKTIIVTRFLKLPDILAIVQRKKXGESE 1554  
QY 1663 QAINRQTALYTLKLLCKNFGAENPPFPVPLTAVKLIAPERKEEKVNLGSLLCIAEVT 1722  
DB 1555 QAINRQTALYTLKLLCKNFGAENPPFPVPLTAVKLIAPERKEEKVNLGSLLCIAEVT 1614

QY 1723 STLEALAIQPLSLMPSLLTTMKNTSELVSSEVYLSALAAQKVVTLPHPFISPYLEGI 1782  
DB 1615 STLEALAIQPLSLMPSLLTTMKNTSELVSSEVYLSALAAQKVVTLPHPFISPYLEGI 1674  
QY 1783 LSQVHLEKITSEMGSASQANIRLTSKKTATTLAPRVLIIPAIKKTYYQIEKNWNHMG 1842  
DB 1675 LSQVHLEKITSEMGSASQANIRLTSKKTATTLAPRVLIIPAIKKTYYQIEKNWNHMG 1734  
QY 1843 PFMSILQEHIGMKKEELTSHQSQTAFLEALDFRAQHSNDLEEVGKTENCIIIDCLVA 1902  
DB 1735 PFMSILQEHIGMKKEELTSHQSQTAFLEALDFRAQHSNDLEEVGKTENCIIIDCLVA 1794  
QY 1903 MVVKLSEVTFRFLPFKLFDMAKTEDAPKORLITTFYNLADCIABKLGKGLFTLFAGHLVKPF 1962  
DB 1795 MVVKLSEVTFRFLPFKLFDMAKTEDAPKORLITTFYNLADCIABKLGKGLFTLFAGHLVKPF 1854  
QY 1963 ADTLQOVNISKTDEAFFDSENDEPKCCLLQFLINCLYKIFLFDTOHFIISKERAXALMMP 2022  
DB 1855 ADTLQOVNISKTDEAFFDSENDEPKCCLLQFLINCLYKIFLFDTOHFIISKERAXALMMP 1914  
QY 2023 LVDQLENRLGSGEEKFQERVTKHLIPCIAQFSVAMADDSLWKPLNYQILLKTRDSSPKVRF 2082  
DB 1915 LVDQLENRLGSGEEKFQERVTKHLIPCIAQFSVAMADDSLWKPLNYQILLKTRDSSPKVRF 1974  
QY 2083 AALITVIALABKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGSPLOS 2142  
DB 1975 AALITVIALABKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGSPLOS 2034  
QY 2143 YF 2144  
DB 2035 YF 2036

## RESULT 8

ABG15232  
ID ABG15232 standard; protein; 1569 AA.  
XX  
AC ABG15232;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #15223.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.  
XX  
FN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS79419.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 45591; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (ii). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (i) is useful in gene therapy techniques to restore normal  
CC activity of (ii) or to treat disease states involving (ii). (ii) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (ii) and its binding partners are useful in medical imaging  
CC of sites expressing (ii). (i) and (ii) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABC00010-ABC0377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 1569 AA;

Query Match 71.8%; Score 7756.5; DB 4; Length 1569;  
Best Local Similarity 93.1%; Pred. No. 0;  
Matches 1553; Conservative 1; Mismatches 1; Indels 113; Gaps 3;  
QY 1 MTLAQQQLRALPOSASLLSRDEVASLLFPDKEAATIDRTAFAIGCTGLEELGIDP 60  
DB 2 MTLAQQQLRALPOSASLLSRDEVASLLFPDKEAATIDRTAFAIGCTGLEELGIDP 61  
QY 61 SFEQPEAPLPSLAKTLERSVOTKAVNKOLDENISLFLHSPYLLKPAQCLBWLTHR 120  
DB 62 SFEQPEAPLPSLAKTLERSVOTKAVNKOLDENISLFLHSPYLLKPAQCLBWLTHR 121  
QY 121 FHIHLYNQDSLACVLPYHETRI FVRVIOQLKINNSKRWFWLLPVKQSGVPLAKGTLLT 180  
DB 122 FHIHLYNQDSLACVLPYHETRI FVRVIOQLKINNSKRWFWLLPVKQSGVPLAKGTLLT 181  
QY 181 HCYKDLGFMDFICSLVTKSVKVAEPGSSAQLRVLLAFYASTIVSALVAADSDNIIA 240  
DB 182 HCYKDLGFMDFICSLVTKSVKVAEPGSSAQLRVLLAFYASTIVSALVAADSDNIIA 241  
QY 241 KLFPYIOKGLKSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKDG 300  
DB 242 KLFPYIOKGLKSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKDG 301  
QY 301 LSCLI VLLQRPESLGKK-----PPFHL CNVPDLITLHGISETYDVSPLL 347  
DB 302 LSCLI VLLQRPESLGKKYVQLILRNGASQRPFPFHL CNVPDLITLHGISETYDVSPLL 361  
QY 348 RYMLPHLVSI IHHVTGEETEGMDGQIYKRHLEAILTKI SLKNLDHLLASLLFEEYISY 407  
DB 362 HYMLPHLVSI IHHVTGEETEGMDGQIYKRHLEAILTKI SLKNLDHLLASLLFEEYISY 421  
QY 408 SSOEEMDSNKVSLNEQFLPLRLLESKYPTLDDVLEHLKEIADLKKOELPHQFVLSL 467  
DB 422 SSOEEMDSNKVSLNEQFLPLRLLESKYPTLDDVLEHLKEIADLKKOELPHQFVLSL 481  
QY 468 TSGGKYQFLADSDTSLMLSLNHPLAPVRILAMNHKKIMKTSKEGYDESGFIKEAVILARLG 527  
DB 482 TSGGKYQ-----EGVDESFIKEAVILARLG 505  
QY 528 DDNIDVLSAISAFETKEHFSSEVTISNLLNLFORAEISKNGEWEVLKIAADIIILKEE 587  
DB 506 DDNIDVLSAISAFETKEHFSSEVTISNLLNLFORAEISKNGEWEVLKIAADIIILKEE 565  
QY 588 ILSENDQLSNQVVVCLLPFVWINDDTESAEMKAIYLSKSGICSLHPLILRGWEALENV 647  
DB 566 ILSENDQLSNQVVVCLLPFVWINDDTESAEMKAIYLSKSGICSLHPLILRGWEALENV 625  
QY 648 IKSTKPKGLIGVANQKMIELADNINILGDPSSMLKNVEDLISVGEBESFNLKQVTFHVI 707  
DB 626 IKSTKPKGLIGVANQKMIELADNINILGDPSSMLKN----- 661

QY 708 LSVLVSCSSSLKETHFPFAIRVFSLLQKKIKKLSVITAVEIPSEWHLEMLDRGIPVEL 767  
DB 662 -----BIPSEWHLEMLDRGIPVEL 681  
QY 768 WAHYVEELNSTORVAVEDSVFLSKKFIYALKAPKSPFKGDIWNWPEQLKEDSRDYLH 827  
DB 682 WAHYVEELNSTORVAVEDSVFLSKKFIYALKAPKSPFKGDIWNWPEQLKEDSRDYLH 741  
QY 828 LLIGLFEMMLNGADAHPRVLMKLFIKVHLEBDVOLFQKFCSVLWYTYGSSLNPLNCVKT 887  
DB 742 LLIGLFEMMLNGADAHPRVLMKLFIKVHLEBDVOLFQKFCSVLWYTYGSSLNPLNCVKT 801  
QY 888 VLQTOALYVGCAMLSQKTOCKHOLASTSSPVVTSLLNLGSPVKEVRRAAIQCLOALSG 947  
DB 802 VLQTOALYVGCAMLSQKTOCKHOLASTSSPVVTSLLNLGSPVKEVRRAAIQCLOALSG 861  
QY 948 VASPYLLIIDLHLSKAEIITSDAAVYIQLDNLATLFEELQREKKLSHQKLSSETLKNLLSCV 1007  
DB 862 VASPYLLIIDLHLSKAEIITSDAAVYIQLDNLATLFEELQREKKLSHQKLSSETLKNLLSCV 921  
QY 1008 YSCPSYIAKDLMKVLQGVNGEMVLSQLLPMAEQLLEKIQKEPTAVLKDAMVHLTLGKY 1067  
DB 922 YSCPSYIAKDLMKVLQGVNGEMVLSQLLPMAEQLLEKIQKEPTAVLKDAMVHLTLGKY 981  
QY 1068 NEFSVSLINEDPKSLDIFIKAHVHTTKELYAGWPTTIQITALEKITKPPFAAISDEKVOQKL 1127  
DB 982 NEFSVSLINEDPKSLDIFIKAHVHTTKELYAGWPTTIQITALEKITKPPFAAISDEKVOQKL 1041  
QY 1128 LRMLFDLLVNCNKSCHCAQTVSVFKGISVNAEQVRIELEPPDKAKPLGTVQQRKQMQ 1187  
DB 1042 LRMLFDLLVNCNKSCHCAQTVSVFKGISVNAEQVRIELEPPDKAKPLGTVQQRKQMQ 1101  
QY 1188 KKSQDLESVQGVGGYQWQVTLIILELQHKKLRSPQILVPTLFLNLSRCLPELPQEQGN 1247  
DB 1102 KKSQDLESVQGVGGYQWQVTLIILELQHKKLRSPQILVPTLFLNLSRCLPELPQEQGN 1161  
QY 1248 MEYTKQLILSCLINICQKLSPDGGKI PKDILDEEFNVELIVOCIRLSEMPQTHHALLL 1307  
DB 1162 MEYTKQLILSCLINICQKLSPDGGKI PKDILDEEFNVELIVOCIRLSEMPQTHHALLL 1221  
QY 1308 LGTVAGIIPDKVLHNINSIFTFMGANVRLDDTYTSFQVINKTVKVIIPALIQSDSGDSIE 1367  
DB 1222 LGTVAGIIPDKVLHNINSIFTFMGANVRLDDTYTSFQVINKTVKVIIPALIQSDSGDSIE 1281  
QY 1368 VSRNVEEIVVKIISVFDALPHVPEHRLPIVLQVLDTLGAEKFLWILLILLFEQVYTKT 1427  
DB 1282 VSRNVEEIVVKIISVFDALPHVPEHRLPIVLQVLDTLGAEKFLWILLILLFEQVYTKT 1341  
QY 1428 VLAAYGEKDALEADTFWPSVCEFSVOHOIQSLMNLQVLLKLPPEKEETIPKAVSF 1487  
DB 1342 VLAAYGEKDALEADTFWPSVCEFSVOHOIQSLMNLQVLLKLPPEKEETIPKAVSF 1401  
QY 1488 NKSESQEEMLQVFNVEHTHSTKOLRHFKFLSVFMSQLSSNNFLKVVESGGPEILKGL 1547  
DB 1402 NKSESQEEMLQVFNVEHTHSTKOLRHFKFLSVFMSQLSSNNFLKVVESGGPEILKGL 1461  
QY 1548 ERLLTETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDLDLKNVALLPTETETFPIVIRGL 1607  
DB 1462 ERLLTETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDLDLKNVALLPTETETFPIVIRGL 1521  
QY 1608 VGNPLPSVRRKALDILLNNKLOQNISWKKTIIVTRFLKLVDPDLIAIVORK 1655  
DB 1522 VGNPLPSVRRKALDILLNNKLOQNISWKKTIIVTRFLKLVDPDLIAIVORK 1569

RESULT 9

ADQ66400

ID ADQ66400 standard; protein; 1229 AA.

XX ADQ66400;

XX DT 07-OCT-2004 (first entry)



XX novel gene; novel protein; tissue marker; molecular weight marker;  
 KW chromosome marker; genetic disorder.  
 XX Unidentified.  
 OS WO2003054152-A2.  
 PN 03-JUL-2003.  
 XX 10-DEC-2002; 2002WO-US039555.  
 XX 10-DEC-2001; 2001US-0339739P.  
 PR 11-DEC-2001; 2001US-0339453P.  
 PR 14-MAR-2002; 2002US-0365091P.  
 PR 14-MAR-2002; 2002US-0365384P.  
 PR 12-APR-2002; 2002US-0372381P.  
 PR 12-APR-2002; 2002US-0372615P.  
 PR 22-APR-2002; 2002US-00128558.  
 PR 24-APR-2002; 2002US-0376049P.  
 XX (HYSE-) HYSEQ INC.  
 XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;  
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;  
 XX WPI; 2003-569235/53.  
 DR N-PSDB; ADE07101.  
 XX New polynucleotides, useful for expressing recombinant proteins for  
 PT analysis, characterization or therapeutic use, or as markers for tissues  
 PT in which the corresponding protein is preferentially expressed.  
 XX Claim 20; SEQ ID NO 1078; 1177pp; English.  
 XX The invention comprises the amino acid and coding sequences of novel  
 CC proteins. The DNA and protein sequences of the invention are useful as:  
 CC markers for tissues in which the corresponding protein is preferentially  
 CC expressed; as molecular weight markers on gels; as chromosome markers or  
 CC tags; to identify chromosomes or to map related gene positions; and to  
 CC compare with endogenous DNA sequences in patients to identify potential  
 CC genetic disorders. The present amino acid sequence represents a protein  
 CC of the invention.  
 XX Sequence 1149 AA;  
 SQ  
 Query Match 46.5%; Score 5026.5; DB 7; Length 1149;  
 Best Local Similarity 52.4%; Pred. No. 0;  
 Matches 1123; Conservative 7; Mismatches 19; Indels 995; Gaps 14;  
 QY 1 MTSLAQQLQRLALPQSDASILSRDEVASLLFDPKAAATIDRDTPAFAGCTGLLEELGIDP 60  
 DB 1 MTSLAQQLQRLALPQSDASILSRDEVASLLFDPKAAATIDRDTPAFAGCTGLLEELGIDP 60  
 QY 61 SFEQFEAPLFSQAKTLERSVQTKAVNKQDENISLFLHLSPYFLLPKPAQKCLEWLIHR 120  
 DB 61 SFEQFEAPLFSQAKTLERSVQTKA----- 85  
 QY 121 FTHILYNQDSLAIACVLPYHETRIFFVRVIQLKKINNSKRWFWLLPVKQSGVPLAKGTLLT 180  
 DB 86 -----QSGVPLAKGTLLT 98  
 QY 181 HCYKDLGFMDFICSLVTKSVKVFAPYGGSSAQLRVLLAFYASTIVSALVAEDVSNIIA 240  
 DB 99 HCYKDLGFMDFICSLVTKSVKVFAPYGGSSAQLRVLLAFYASTIVSALVAEDVSNIIA 158  
 QY 241 KLFPYQKGLKSSLPDYRAATYMLICOISVKVTMENTFVNSLASOIKITLTIPSLIKOG 300  
 DB 159 KLFPYIQK-----IITLTIPSLIKOG 181  
 QY 301 LSCILVLLQKQKPSLGKPPHLCNVPDLITILHGISETYDVSPLRLRYMLPHLWVSIH 360  
 DB 182 LSCILVLLQKQKPSLGKK----- 200  
 QY 361 HVTGEETEGMDGQIYKRHLRAITKISLKNLHLLASLLFEEYISYSSQEMDSNKVSL 420  
 DB 201 ----- 200  
 QY 421 LNEQFLPLIRLLESKYPRTLDDVVLEBHLKETIADLKQELFHQFVLSLSTGGKYOFLADSD 480  
 DB 201 ----- 200  
 QY 481 TSLMSLNHPLAPVRILAMNHLKKIMKTSKEGVDSFIKEAVLARLGGDNIDVLSAISA 540  
 DB 201 ----- 200  
 QY 541 FEIPEHFSSEVTISNLLNLFQRAELSKNGEWYEVLIKIAADILIKEEILSENDQLSNOVV 600  
 DB 201 -----YEVLKIAADILIKEEILSENDQLSNOVV 228  
 QY 601 VCLLPFVIVNDDTESAEMKIAIYLSKSGICSLHPLLRGWEALENVIKSTKPGKLI GVA 660  
 DB 229 VCLLPFVIVNDDTESAEMKIAIYLSKSGICSLHPLLRGWE----- 270  
 QY 661 NQMIELADNINLGDPSMLKMVEDLISVGBEESFNLKQKTFHVLISVLVSCSSLSKE 720  
 DB 271 ----- 270  
 QY 721 THFFPAIRVFSLLQKKIKKLESVITAVEIPSEWHIELMLDRGIPVELWAHYVEELNSTQR 780  
 DB 271 ----- 270  
 QY 781 VAVEDSVFLVFLSKFIYALKAPKSPFGKDIWNPEQLKEDSRDYLHLLIGLFEMWLNKA 840  
 DB 271 -----GDIWNPEQLKEDSRDYLHLLIGLFEMWLNKA 302  
 QY 841 DAVHFRVLMKFIKVHLEDFQLFKFCVLTWYTGSSLSNPLNCSVKTVLQOALYVGCAM 900  
 DB 303 DAVHFRVLMKLFIK----- 316  
 QY 901 LSSQKTOCKHQLASISSPVVTSLLINLSPVKVEVRAAICQICLQALSGVASPFYLIIDHLI 960  
 DB 317 ----- 316  
 QY 961 SKAEBITSDAAVVIQDLATLFELOREKKLKHQKLSKSETLKNLSCVYSCPSYIAKOLMK 1020  
 DB 317 ----- 316  
 QY 1021 VLQGVNGEMVLSQLLPMAEQLEKIQEPTAVLKDEAMVHLTLTKYNEFSVSLNEDPK 1080  
 DB 317 -----MVLSQLLPMAEQLEKIQEPTAVLKDEAMVHLTLTKYNEFSVSLNEDPK 368  
 QY 1081 SLDIFIKAVHTTKELYAGMPTIQTITALEKIKTPFFAAISDEKVOQKLLRMLFDLLVNCN 1140  
 DB 369 SLDIFIKAVHTTKELYAGMPTIQTITALEKILRIY-----ISD----- 405  
 QY 1141 SHCAQTSSVSVFKGISVNAEQVRIELEPPDKAPLGTVOQKRRKQKMOQKSQDLESVOEVG 1200  
 DB 406 ----- 405  
 QY 1201 GSYQWRVTLILELLOHKKKLRSPQIIVPTLFNLLSRCLEPLPQEQGNMEYTKQILISCLL 1260  
 DB 406 -----PAL-----GNNR-----CLL 415  
 QY 1261 NTCOKLSPGGKIPKDIIDEEKFNVELIVQCIRLSMPOTHHALLLGTVAGIFPDKVL 1320  
 DB 416 NTCOKLSPGGKIPKDIIDEEKFNVELIVQCIRLSMPOTHHALLLGTVAGIFPDKVL 475  
 QY 1321 HNIMSIFFTWGANVWELDDTYSFQVINKTVKMWIPALIQSDSGDSIEVSRNVEEIVVKII 1380  
 DB 476 HNIMSIFFTWGANVWELDDTYSFQVINKTVKMWIPALIQSDSGDSIEVSRNVEEIVVKII 535  
 QY 1381 SVFVDALPHVPEHRRLLPILVOLVDTLGAEKFLWILLILPEQVYVTKVLAAYAGEKDAIL 1440  
 DB 536 SVFVDALPHVPEHRRLLPILVOLVDTLGAEKFLWILLILPEQVYVTKVLAAYAGEKDAIL 595





Db 481 LAIPQLPSLMPSLTMTKNTSELVSSEVYLLSALAALQKVVETLPHFISPYLEGILSQVI 540  
QY 1788 HLEKITSMSGASQANIRLTSKKTATTLAPRVLLPAIKTTYKQIEKNWKNHMGPPMSI 1847  
Db 541 HLEKITSMSGASQANIRLTSKKTATTLAPRVLLPAIKTTYKQIEKNWKNHMGPPMSI 600  
QY 1848 LOEHIGXMKKEBELTSHQSQTAFLEALDFRAHSENDLEEVGKTENCIIIDCLVAMVVKL 1907  
Db 601 LOEHIGVWKKEBELTSHQSQTAFLEALDFRAHSENDLEEVGKTENCIIIDCLVAMVVKL 660  
QY 1908 SEVTRPLPFKLFDAKTEDAPKDRLLTFYNLADCAIEKLGKLTFLFAGHLVKFPADTLX 1967  
Db 661 SEVTRPLPFKLFDAKTEDAPKDRLLTFYNLADCAIEKLGKLTFLFAGHLVKFPADTLN 720  
QY 1968 QVNISKTDEAFDSENDPEKCLLQFILNCLYKIFLFDTHQHFISKERAXALMPLVDQL 2027  
Db 721 QVNISKTDEAFDSENDPEKCLLQFILNCLYKIFLFDTHQHFISKERAXALMPLVDQL 780  
QY 2028 ENRLGGEKFOBRVTKHLIPCTIAQFSVAMADSLWKPLNYQILLKTRDSSPKRPAALIT 2087  
Db 781 ENRLGGEKFOBRVTKHLIPCTIAQFSVAMADSLWKPLNYQILLKTRDSSPKRPAALIT 840  
QY 2088 VLALAEKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLOSYF 2144  
Db 841 VLALAEKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLOSYF 897  
  
RESULT 12  
ADQ96218  
ID ADQ96218 standard; protein; 897 AA.  
XX  
AC ADQ96218;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
DE T cell activation associated protein #198.  
XX  
KW antiallergic; antiarthritic; antiasthmatic; antidiabetic; anti-HIV;  
KW antimicrobial; antirheumatic; immunosuppressive; neuroprotective;  
KW gene therapy; T cell activation; diagnosis; autoimmune disease;  
KW rheumatoid arthritis; asthma; multiple sclerosis; diabetes;  
KW allergic disease; infectious disease; AIDS; chronic rejection; organ;  
KW bone-marrow transplant.  
XX  
OS Homo sapiens.  
XX  
PN WO2004058805-A2.  
XX  
XX 15-JUL-2004.  
XX  
XX 25-DEC-2003; 2003WO-JP016715.  
XX  
XX 26-DEC-2002; 2002JP-00376365.  
XX  
XX 27-DEC-2002; 2002US-0436473P.  
XX  
XX 25-APR-2003; 2003JP-00122113.  
XX  
XX 28-APR-2003; 2003US-0465792P.  
XX  
XX 21-OCT-2003; 2003JP-00360559.  
XX  
XX 22-OCT-2003; 2003US-0512846P.  
XX  
XX (ASAHI-) ASAHI KASEI PHARMA CORP.  
XX  
XX Matsuda A, Yoneta S;  
XX  
XX WPI; 2004-593134/57.  
XX  
XX N-PSDB; ADQ96217.  
XX  
XX New purified protein involved in T cell activation, useful for  
XX diagnosing, preventing and/or treating acquired immunodeficiency  
XX syndrome, autoimmune (e.g. rheumatoid arthritis, and diabetes), allergic  
XX and infectious diseases.  
XX  
XX Claim 1; SEQ ID NO 396; 2028pp; English.  
XX  
XX

CC The invention relates to purified proteins and genes encoding them, that  
CC are involved in T cell activation (I) and has an amino acid deletion,  
CC substitution or addition in the amino acid sequences. The methods and  
CC compositions of the present invention are useful for the diagnosis,  
CC prevention and/or treatment of autoimmune disease (rheumatoid arthritis,  
CC asthma, multiple sclerosis and diabetes), allergic disease, infectious  
CC disease, AIDS, and acute or chronic rejection at organ transplant or bone  
CC marrow transplant. This sequence corresponds to a protein involved in T  
CC cell activation.  
XX

XX SQ Sequence 897 AA;

Query March 41.6%; Score 4498; DB 8; Length 897;

Best Local Similarity 99.2%; Pred. No. 8.7e-318;

Matches 890; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1248 MEYTKQILSLCLLNICQKLSPDGKIPKDIIDDEKFNVELIVQCIRLSEMPQTHHALLL 1307

Db 1 MEYTKQILSLCLLNICQKLSPDGKIPKDIIDDEKFNVELIVQCIRLSEMPQTHHALLL 60

QY 1308 LGTVAGIPDPKVLNINMISIFTFMGANVRLDDTYSFQVINKTVKVIIPALIQSDSGDSIE 1367

Db 61 LGTVAGIPDPKVLNINMISIFTFMGANVRLDDTYSFQVINKTVKVIIPALIQSDSGDSIE 120

QY 1368 VSRNVEEIVWKIISVFVDALPHVPEHRLPILVOLVDTLGAEKFLWILLILFEQVYTKT 1427

Db 121 VSRNVEEIVWKIISVFVDALPHVPEHRLPILVOLVDTLGAEKFLWILLILFEQVYTKT 180

QY 1428 VLAAYGEKDAILEADTEFWFSVCEFSVQHQIQSLMNILOYLKLPEKEETIPKAVSF 1487

Db 181 VLAAYGEKDAILEADTEFWFSVCEFSVQHQIQSLMNILOYLKLPEKEETIPKAVSF 240

QY 1488 NKSEQSEMLQVFNVEHTSKQLRHFKFLSVFMSQLLSSNLFKKVSGGPELKGLE 1547

Db 241 NKSEQSEMLQVFNVEHTSKQLRHFKFLSVFMSQLLSSNLFKKVSGGPELKGLE 300

QY 1548 ERLETVLGYISAVAQSMERNADKLTVKFRALLSKAYDLDLKVALLPTETFIPIVIRGL 1607

Db 301 ERLETVLGYISAVAQSMERNADKLTVKFRALLSKAYDLDLKVALLPTETFIPIVIRGL 360

QY 1608 VGNPLPSVRRKALDILNNKLQONISWKKTIIVTRFLKLPVLLAIIVORKKKEGEEQAINR 1667

Db 361 VGNPLPSVRRKALDILNNKLQONISWKKTIIVTRFLKLPVLLAIIVORKKKEGEEQAINR 420

QY 1668 QATALYTLKLLCKNFGAENPDPPVPLSTAVKLIAPERKEEKVIGSALLCIAEVTSTLEA 1727

Db 421 QATALYTLKLLCKNFGAENPDPPVPLSTAVKLIAPERKEEKVIGSALLCIAEVTSTLEA 480

QY 1728 LAIPQLPSLMPSLTMTKNTSELVSSEVYLLSALAALQKVVETLPHFISPYLEGILSQVI 1787

Db 481 LAIPQLPSLMPSLTMTKNTSELVSSEVYLLSALAALQKVVETLPHFISPYLEGILSQVI 540

QY 1788 HLEKITSEMGASQANIRLTSKKTATTLAPRVLLPAIKTTYKQIEKNWKNHMGPPMSI 1847

Db 541 HLEKITSEMGASQANIRLTSKKTATTLAPRVLLPAIKTTYKQIEKNWKNHMGPPMSI 600

QY 1848 LOEHIGXMKKEBELTSHQSQTAFLEALDFRAHSENDLEEVGKTENCIIIDCLVAMVVKL 1907

Db 601 LOEHIGVWKKEBELTSHQSQTAFLEALDFRAHSENDLEEVGKTENCIIIDCLVAMVVKL 660

QY 1908 SEVTRPLPFKLFDAKTEDAPKDRLLTFYNLADCAIEKLGKLTFLFAGHLVKFPADTLX 1967

Db 661 SEVTRPLPFKLFDAKTEDAPKDRLLTFYNLADCAIEKLGKLTFLFAGHLVKFPADTLN 720

QY 1968 QVNISKTDEAFDSENDPEKCLLQFILNCLYKIFLFDTHQHFISKERAXALMPLVDQL 2027

Db 721 QVNISKTDEAFDSENDPEKCLLQFILNCLYKIFLFDTHQHFISKERAXALMPLVDQL 780

QY 2028 ENRLGGEKFOBRVTKHLIPCTIAQFSVAMADSLWKPLNYQILLKTRDSSPKRPAALIT 2087

Db 781 ENRLGGEKFOBRVTKHLIPCTIAQFSVAMADSLWKPLNYQILLKTRDSSPKRPAALIT 840

QY 2088 VLALAEKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLOSYF 2144

```

Db      841 VLALAEKLNKENVYLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLQSYF 897
|||||
RESULT 13
ADQ96152
ID ADQ96152 standard; protein; 897 AA.
XX AC ADQ96152;
XX DT 07-OCT-2004 (first entry)
XX DE T cell activation associated protein #165.
XX KW antiallergic; antiarthritic; antiasthmatic; antidiabetic; anti-HIV;
KW antimicrobial; antirheumatic; immunosuppressive; neuroprotective;
KW gene therapy; T cell activation; diagnosis; autoimmune disease;
KW rheumatoid arthritis; asthma; multiple sclerosis; diabetes;
KW allergic disease; infectious disease; AIDS; chronic rejection; organ;
KW bone-marrow transplant.
XX OS Homo sapiens.
XX PN WO2004058805-A2.
XX PD 15-JUL-2004.
XX PF 25-DEC-2003; 2003WO-JP016715.
XX PR 26-DEC-2002; 2002JP-00376365.
XX PR 27-DEC-2002; 2002US-0436473P.
XX PR 25-APR-2003; 2003JP-00122113.
XX PR 28-APR-2003; 2003US-0465792P.
XX PR 21-OCT-2003; 2003JP-00360559.
XX PR 22-OCT-2003; 2003US-0512846P.
XX PA (ASAH-) ASahi Kasei Pharma Corp.
XX PI Matsuda A, Yoneta S;
XX WPI; 2004-593134/57.
XX DR N-PSDB; ADQ96151.
XX PT New purified protein involved in T cell activation, useful for
PT diagnosing, preventing and/or treating acquired immunodeficiency
PT syndrome, autoimmune (e.g. rheumatoid arthritis, and diabetes), allergic
PT and infectious diseases.
XX PS Claim 1; SEQ ID NO 330; 2828pp; English.
XX CC The invention relates to purified proteins and genes encoding them, that
CC are involved in T cell activation (I) and has an amino acid deletion,
CC substitution or addition in the amino acid sequences. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of autoimmune disease (rheumatoid arthritis,
CC asthma, multiple sclerosis and diabetes), allergic disease, infectious
CC disease, AIDS, and acute or chronic rejection at organ transplant or bone
CC marrow transplant. This sequence corresponds to a protein involved in T
CC cell activation.
XX SQ Sequence 897 AA;

Query Match 41.6%; Score 4494; DB 8; Length 897;
Best Local Similarity 99.2%; Pred. No. 1.7e-317;
Matches 890; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1248 MEYTKQLILSCLLNKCOKLSPDGGKIPKIDILDEEFNVELIVQCIRLSEMPQTHHALLL 1307
DB 1 MEYTKQLILSCLLNKCOKLSPDGGKIPKIDILDEEFNVELIVQCIRLSEMPQTHHALLL 60

QY 1308 LGTVAGIFPDVKLVHNMISFTFMGANVRLDDTYSFQVINKTVKMWIPALIQSDSGDSIE 1367
DB 61 LGTVAGIFPDVKLVHNMISFTFMGANVRLDDTYSFQVINKTVKMWIPALIQSDSGDSIE 120

```

```

QY 1368 VSRNVEEIVVKIISVFDALPHVPEHRRLPILVQLVDTLGAEKFLWILLILLPEQVYTKT 1427
DB 121 VSRNVEEIVVKIISVFDALPHVPEHRRLPILVQLVDTLGAEKFLWILLILLPEQVYTKT 180
QY 1428 VLAAYGEKDAILEADTEFWFVSCCEFSVQHQIQSLMNIQYLLKLPKKEETIPKAVSF 1487
DB 181 VLAAYGEKDAILEADTEFWFVSCCEFSVQHQIQSLMNIQYLLKLPKKEETIPKAVSF 240
QY 1488 NKSESOEMLQVFNVEHTSKOLRHFPLSVFMSOLLSSNPNFKKVVSGGPEILLKGLE 1547
DB 241 NKSESOEMLQVFNVEHTSKOLRHFPLSVFMSOLLSSNPNFKKVVSGGPEILLKGLE 300
QY 1548 ERLLETVLGYISAVAQSMERNADKLVKFWRALLSKAYDLDLKVALLTETETIPVIRGL 1607
DB 301 ERLLETVLGYISAVAQSMERNADKLVKFWRALLSKAYDLDLKVALLTETETIPVIRGL 360
QY 1608 VGNPLPSVRKALDILNNKLQONISWKKITVTRFLKVPDILAIVQRKKKEGEEQAIRN 1667
DB 361 VGNPLPSVRKALDILNNKLQONISWKKITVTRFLKVPDILAIVQRKKKEGEEQAIRN 420
QY 1668 QTALYTLKLLCKNFGAENPDPFVPLXTAVKLIAPERKEKNVLSALLCIAEVTSTLEA 1727
DB 421 QTALYTLKLLCKNFGAENPDPFVPLXTAVKLIAPERKEKNVLSALLCIAEVTSTLEA 480
QY 1728 LAIPOLPSLMPSLTMTKNTSELVSEVYLLSALAALQKVETLPHFISPYLEGILSQVI 1787
DB 481 LAIPOLPSLMPSLTMTKNTSELVSEVYLLSALAALQKVETLPHFISPYLEGILSQVI 540
QY 1788 HLEKITSEMGASQANIRLTSKKTLATLAPRVLLPAIKKTYKQIEKNWKNHMGPFMSI 1847
DB 541 HLEKITSEMGASQANIRLTSKKTLATLAPRVLLPAIKKTYKQIEKNWKNHMGPFMSI 600
QY 1848 LOEHIGKMKKEELTSHQSOLTAFFLEALDFRAQHSNDLEEVGKTENCIIICLVAMVVKL 1907
DB 601 LOEHIGKMKKEELTSHQSOLTAFFLEALDFRAQHSNDLEEVGKTENCIIICLVAMVVKL 660
QY 1908 SEVTFRPLFFKLFDWAKTEDAPKDRLLTFYNLADCIAEKLGLFTLFAGHLVKPFADTLX 1967
DB 661 SEVTFRPLFFKLFDWAKTEDAPKDRLLTFYNLADCIAEKLGLFTLFAGHLVKPFADTLX 720
QY 1968 QVNISKTDPAFPDSENDPEKCLLILQFILNCLYKIFLFDTHFISKERAXALMPLVDQL 2027
DB 721 QVNISKTDPAFPDSENDPEKCLLILQFILNCLYKIFLFDTHFISKERAXALMPLVDQL 780
QY 2028 ENRLGGEKGFQBRVTKHLIPCTAOPSVAMADSLWKPLNYQILLKTRDSSPKVRPAAIT 2087
DB 781 VNRLGGEKGFQBRVTKHLIPCTAOPSVAMADSLWKPLNYQILLKTRDSSPKVRPAAIT 840
QY 2088 VLALAEKLNKENVYLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLQSYF 2144
DB 841 VLALAEKLNKENVYLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLQSYF 897

RESULT 14
ADQ04615
ID ADQ04615 standard; protein; 734 AA.
XX AC ADQ04615;
XX DT 20-MAY-2004 (first entry)
XX DE Human protein of the invention SEQ ID NO:3300.
XX KW human; gene therapy; diagnostic marker; pharmaceutical.
XX OS Homo sapiens.
XX PN EP1347046-A1.
XX PD 24-SEP-2003.
XX PF 12-APR-2002; 2002EP-00008400.

```

XX 22-MAR-2002; 2002JP-00137785.  
XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yanamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahazi K, Masuho Y;  
XX WPI; 2003-723558/69.  
DR N-PSDB; ADM02172.  
XX  
PT New polynucleotides and polypeptides are useful in gene therapy, for  
PT developing a diagnostic marker or medicines for regulating their  
PT expression and activity, or as a target of gene therapy.  
XX  
PS Claim 1; SEQ ID NO 3300; 305pp; English.  
XX  
CC The invention relates to a novel human polynucleotide and the encoded  
CC polypeptide. A polynucleotide of the invention may have a use in gene  
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful  
CC as a primer for synthesizing the polynucleotide or as a probe for  
CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are  
CC useful in gene therapy, for developing a diagnostic marker or medicines  
CC for regulating their expression and activity, or as a target of gene  
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides  
CC are useful as pharmaceutical agents. The present sequence represents a  
CC protein sequence of the invention.  
XX  
SQ Sequence 734 AA;  
Query Match 34.2%; Score 3694; DB 7; Length 734;  
Best Local Similarity 100.0%; Pred. No. 1.7e-259;  
Matches 734; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTSLAQQLQRLALPQSDASLLSRDEVASLLFPDKAATIDRTDTAFAGTGTGEEELGIDP 60  
DB 1 MTSLAQQLQRLALPQSDASLLSRDEVASLLFPDKAATIDRTDTAFAGTGTGEEELGIDP 60  
QY 61 SFEQEPAPLFSOLAKTLERSVQTKAVNKOLDENISLFLHLSPYLLKPAQKCLEWLIHR 120  
DB 61 SFEQEPAPLFSOLAKTLERSVQTKAVNKOLDENISLFLHLSPYLLKPAQKCLEWLIHR 120  
QY 121 FHIHLNQDSLACVLPYHETRFVRIQLLKINNSKRWFWLLPVKQSGVPLAKGTLLT 180  
DB 121 FHIHLNQDSLACVLPYHETRFVRIQLLKINNSKRWFWLLPVKQSGVPLAKGTLLT 180  
QY 181 HCYKDLGFMDFICSLVTKSVKVPFAEPYGSQAQLRVLLAFYASTIVSALVAEEDVSDNIIA 240  
DB 181 HCYKDLGFMDFICSLVTKSVKVPFAEPYGSQAQLRVLLAFYASTIVSALVAEEDVSDNIIA 240  
QY 241 KLPPYIQKGLKSLPYRAATYMIICQISVKVTMENTFVNSLASQIKTLTKIPSLIKDG 300  
DB 241 KLPPYIQKGLKSLPYRAATYMIICQISVKVTMENTFVNSLASQIKTLTKIPSLIKDG 300  
QY 301 LSLCLVLLQKQKESLGKPKFPFHLGNVPLDITLHIGSETYDVSPLLRYMLPHLVVSIH 360  
DB 301 LSLCLVLLQKQKESLGKPKFPFHLGNVPLDITLHIGSETYDVSPLLRYMLPHLVVSIH 360  
QY 361 HVTGEETEGMDGQIQYKRLHLEAITKISLKNLPHLLASLLFEBYIYSYSGEEMDSKNVSL 420  
DB 361 HVTGEETEGMDGQIQYKRLHLEAITKISLKNLPHLLASLLFEBYIYSYSGEEMDSKNVSL 420  
QY 421 LNEQFLPLIRLLESKYPTLDVLEHLKEIADLKQELPHQFVSLSTSGKQYFLADSD 480  
DB 421 LNEQFLPLIRLLESKYPTLDVLEHLKEIADLKQELPHQFVSLSTSGKQYFLADSD 480  
QY 481 TSLMLSLNHPLAPVRILAMNHLKIMKTSKEGVDSFKEAVLARLGDDNIDVLSAISA 540  
DB 481 TSLMLSLNHPLAPVRILAMNHLKIMKTSKEGVDSFKEAVLARLGDDNIDVLSAISA 540  
QY 541 PEIFKHFSEVTSINLNLFORAELSKNGEWYEVLKIAADILIKEILSENDSQV 600

DB 541 PEIFKHFSEVTSINLNLFORAELSKNGEWYEVLKIAADILIKEILSENDSQV 600  
QY 601 VCLLPFFVINNDTSAEMKIAIYLSKSGICSLHPLLRGWEAELENVIKSTPKGLIGVA 660  
DB 601 VCLLPFFVINNDTSAEMKIAIYLSKSGICSLHPLLRGWEAELENVIKSTPKGLIGVA 660  
QY 661 NQKMIELLADNINIGDPSSMLKMWEDLISVGEESFNLKQKVTFFHVLVSVLSCSSLUKE 720  
DB 661 NQKMIELLADNINIGDPSSMLKMWEDLISVGEESFNLKQKVTFFHVLVSVLSCSSLUKE 720  
QY 721 THPPFAIRVFSLQ 734  
DB 721 THPPFAIRVFSLQ 734  
RESULT 15  
AAWS4099  
ID AAWS4099 standard; protein; 515 AA.  
XX  
AC AAWS4099;  
XX  
DT 28-SEP-1998 (first entry)  
XX  
DE Homo sapiens BAP28 sequence.  
XX  
KW BARD1; ring protein; BRCA1; breast cancer; risk; diagnosis.  
XX  
OS Homo sapiens.  
XX  
PN WO9812327-A2.  
XX  
PD 26-MAR-1998.  
XX  
PF 19-SEP-1997; 97WO-US016842.  
XX  
PR 20-SEP-1996; 96US-0025296P.  
PR 03-APR-1997; 97US-0042611P.  
PR 04-APR-1997; 97US-0042985P.  
XX  
PA (TEXA ) UNIV TEXAS SYSTEM.  
XX  
PI Bowcock AM, Baer R;  
XX  
DR WPI; 1998-230317/20.  
DR N-PSDB; AAV24135.  
XX  
PT DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which as  
PT breast cancer antigen, BRCA1, binding proteins are useful to identify  
PT patient having or at risk of developing cancer.  
XX  
PS Disclosure; Page 287-288; 348pp; English.  
XX  
CC The sequence is that of a protein which can be used in the preparation of  
CC the recombinant breast cancer antigen, BRCA1, binding proteins BARD1,  
CC B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a  
CC BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence, specifically  
CC a wild type BARD1 composition for the detection or purification of BRCA1,  
CC useful to identify a patient having, or at risk of developing cancer.  
CC BARD1 can be used in the preparation of an anti-BARD1 antibody, and in  
CC the detection and purification of a BRCA1 protein. BARD1, B123, BE2,  
CC BE14, BE31 or BE445 can be used in the identification of a binding protein  
CC agonist or antagonist that alters the binding of BARD1, B123, BE2, BE14,  
CC BE31 or BE445 to BRCA1 or the biological activity of the BRCA1-BARD1,  
CC B123, BE2, BE14, BE31 or BE445 complex. The antibodies can be used to  
CC detect BARD1, B123, BE2, BE14, BE31 or BE445, a specific anti-BARD1  
CC antibody can be used to identify a patient having or at risk of  
CC developing cancer  
XX  
SQ Sequence 515 AA;  
Query Match 23.9%; Score 2585; DB 2; Length 515;  
Best Local Similarity 99.0%; Pred. No. 5.4e-179;  
Matches 510; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy	1630	NISWKKTIVTRPFLKVPDLLAI	VOR	KKKEBEEBQAINROTAL	YTLKLLCNFGAENPDPP	1689	
Db	1	NISWKKTIVTRPFLKVPDLLAI	VOR	KKKEBEEBQAINROTAL	YTLKLLCNFGAENPDPP	60	
Qy	1690	VPVLXTAVKLIAPERKEKNVLSAL	CIAE	VSTLTLEALAI	POLPSLMPSLLLTKMNTSE	1749	
Db	61	VPVLXTAVKLIAPERKEKNVLSAL	CIAE	VSTLTLEALAI	POLPSLMPSLLLTKMNTSE	120	
Qy	1750	LVSEVYLLSALAAQKVETLPHF	ISPY	LEGILSQVILHEKIT	SEMGSSASQANIRLTSL	1809	
Db	121	LVSEVYLLSALAAQKVETLPHF	ISPY	LEGILSQVILHEKIT	SEMGSSASQANIRLTSL	180	
Qy	1810	KKTLLATTLAPRVLLPAIKKTY	KOT	EKNWKNHMGPFMS	ILQEHICXMKKBELTSHQSOLTA	1869	
Db	181	KKTLLATTLAPRVLLPAIKKTY	KOT	EKNWKNHMGPFMS	ILQEHIGAMKKBELTSHQSOLTA	240	
Qy	1870	FFLEALDFRAQSENDL	EEV	GKTENCII	DCLVAMVVKLSVTRPRPLFFKLF	FDWAKTEDAP	1929
Db	241	FFLEALDFRAQSENDL	EEV	GKTENCII	DCLVAMVVKLSVTRPRPLFFKLF	FDWAKTEDAP	300
Qy	1930	KDRLLTFYNLADCTIAEKL	KGFL	TLFAGHLVKFPAD	TLXQVNI	SKTDEAFDSENDPEKCC	1989
Db	301	KDRLLTFYNLADCTIAEKL	KGFL	TLFAGHLVKFPAD	TLXQVNI	SKTDEAFDSENDPEKCC	360
Qy	1990	LLLOFTLNCILYKIFL	EDTO	HFISKERAXALMPL	VDQLENRLGGBEFQBRVTKHIL	IPCI	2049
Db	361	LLLOFTLNCILYKIFL	EDTO	HFISKERAXALMPL	VDQLENRLGGBEFQBRVTKHIL	IPCI	420
Qy	2050	AQFSVAMADDSLWKPLNYQ	ILL	TKTRUSSPKVR	FAALITVLALAEKUKENYIVLL	PESIPF	2109
Db	421	VQFSVAMADDSLWKPLNYQ	ILL	TKTRUSSPKVR	FAALITVLALAEKUKENYIVLL	PESIPF	480
Qy	2110	LAELMEDECEVEHQCKT	IQO	ETVLGEPLQSYF		2144	
Db	481	LAELMEDECEVEHQCKT	IQO	ETVLGEPLQSYF		515	

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 09:31:57 ; Search time 262 Seconds  
(without alignments)  
4190.450 Million cell updates/sec

Title: US-09-603-665-5  
Perfect score: 10807  
Sequence: 1 MTSLAQQLRLALPQSDASL.....CKTKTQQLTVLGEPLQSYF 2144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10799	99.9	2144	1 BP28_HUMAN	Q9h583 homo sapien
2	5807.5	53.7	2159	2 Q7SY48	Q7sy48 brachydanio
3	5562	51.5	1106	2 Q6P197	Q6p197 homo sapien
4	4729	43.8	958	1 BP28_MACFA	Q9gm44 macaca fasc
5	4494	41.6	897	2 Q8N7L7	Q8n7l7 homo sapien
6	3935.5	36.4	1336	2 Q7T152	Q7t152 brachydanio
7	3669.5	34.0	1278	2 Q7T153	Q7t153 brachydanio
8	3099	28.7	733	2 Q8BLJ4	Q8blj4 mus musculus
9	2094.5	19.4	2098	2 Q7PMD6	Q7pmd6 anopheles g
10	2051.5	19.0	2096	1 BP28_DROME	Q9vm75 drosophila
11	2051.5	19.0	2096	2 Q6AWS0	Q6aws0 drosophila
12	1926	17.8	408	2 Q8CCT5	Q8cct5 mus musculus
13	1779	16.5	349	2 Q96ES5	Q96es5 mus sapien
14	1682	15.6	349	2 Q8VCK1	Q8vck1 mus sapien
15	1462.5	13.5	1690	2 Q8T9E7	Q8t9e7 drosophila
16	1194.5	11.1	1830	1 BP28_ARATH	Q9c8z4 arabidopsis
17	1127	10.4	1857	2 Q6BXQ6	Q6bxq6 debaryomyce
18	1030.5	9.5	1788	2 Q7R2M8	Q7r2m8 neurospora
19	1006.5	9.3	1770	2 Q6FT93	Q6ft93 candida gla
20	1003.5	9.3	1774	2 Q6CJ57	Q6cj57 kluyveromyce
21	990.5	9.2	1649	1 YG44_SCHPO	Q60179 schizosacch
22	956	8.8	1769	1 YJK9_YEAST	P42945 saccharomyc
23	950.5	8.8	1774	2 Q6YJ78	Q75418 ashbya goss
24	903	8.4	1635	2 Q6C457	Q6c457 yarrowia li
25	857	7.9	1650	1 BP28_CAEEL	Q23495 caenorhabdi
26	800	5.6	120	2 Q6P664	Q6p664 homo sapien
27	410.5	3.8	2381	2 Q7R4R8	Q7r4r8 giardia lam
28	365.5	3.4	2574	2 Q6BFG2	Q6bfg2 paramecium
29	320.5	3.0	2736	2 Q6BMQ6	Q6bmq6 debaryomyce
30	317.5	2.9	2670	1 YAQ5_SCHPO	Q10105 schizosacch
31	309.5	2.9	2230	1 G0A4_HUMAN	Q13439 homo sapien

32 305 2.8 2671 2 Q754A3 Q754a3 ashbya goss  
33 295 2.7 5058 2 Q86UQ4 Q86uq4 homo sapien  
34 294 2.7 2792 2 Q814R2 Q814r2 plasmodium  
35 293 2.7 3130 2 Q81DX6 Q81dx6 plasmodium  
36 293 2.7 3254 2 Q7YWE9 Q7ywe9 plasmodium  
37 293 2.7 3254 2 Q9BK45 Q9bk45 plasmodium  
38 292 2.7 3130 2 Q9BK46 Q9bk46 plasmodium  
39 291.5 2.7 3080 2 Q7YWF0 Q7ywf0 plasmodium  
40 291.5 2.7 3081 2 Q7YWF1 Q7ywf1 plasmodium  
41 291.5 2.7 3203 2 Q7YWE6 Q7ywe6 plasmodium  
42 291.5 2.7 3203 2 Q7YWE7 Q7ywe7 plasmodium  
43 290.5 2.7 2866 2 Q6FW99 Q6fw99 candida gla  
44 290 2.7 3096 2 Q7YWF2 Q7ywf2 plasmodium  
45 290 2.7 3256 2 Q7YWE8 Q7ywe8 plasmodium

#### ALIGNMENTS

RESULT 1  
BP28\_HUMAN STANDARD; PRT; 2144 AA.  
AC Q9H5E3; Q9NW23;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Protein BAP28.  
GN Name=BAP28;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND VARIANTS SER-1694; ALA-1854; ASP-1967 AND  
RP GLY-2017.  
RA Bougueret L., Chumakov I., Barry C., Cohen-Akenine A.;  
RT "A novel BAP28 gene and protein."  
RL Patent number WO0100669, 04-JAN-2001.  
RN [2]  
RP SEQUENCE OF 1534-2144 FROM N.A.  
RA Cobley V.,  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RN SEQUENCE OF 1777-2144 FROM N.A.  
RX PubMed=14702039; DOI=10.1038/ngl1285;  
RX Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RX Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RX Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RX Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
RX Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
RX Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
RX Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
RX Omura Y., Abe K., Kamiyama M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
RX Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,  
RX Fujimori K., Tanai H., Kimata S., Watanabe M., Hiraoa K., Chiba Y.,  
RX Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,  
RX Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
RX Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,  
RX Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RX Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
RX Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RX Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RX Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RX Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RX Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RX Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Okamoto S.,  
RX Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RX Okitani R., Kawakami T., Noguchi S., Itoh Y., Shigeta K., Senba T.,  
RX Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RX Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RX Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RX Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
RX Nakai K., Yada T., Nakamura Y., Ohara O., Iisaga T., Sugano S.;



Db 1621 DLNNKQQNISKWKYIVTRFLKLVDPDLAIVQKKKEGEEOAINRQTALYTLKLCN 1680  
Qy 1681 FGAENPDPPVPLVTKLAVKLIAPERKEKNVGLSALLCIAEVTSTLEALAIPLQPLSMPSL 1740  
Db 1681 FGAENPDPPVPLVTKLAVKLIAPERKEKNVGLSALLCIAEVTSTLEALAIPLQPLSMPSL 1740  
Qy 1741 LTTMKNTSELVSSEVILLSSALALQKVETLPHFIPSPYLEGILSQVHLEKITSSEMGSSAS 1800  
Db 1741 LTTMKNTSELVSSEVILLSSALALQKVETLPHFIPSPYLEGILSQVHLEKITSSEMGSSAS 1800  
Qy 1801 QANIRLTSKKTATLTIAPRVLLPAIKTYKQLEKNWKNHGMSPMSTLOSHICXMKKEEL 1860  
Db 1801 QANIRLTSKKTATLTIAPRVLLPAIKTYKQLEKNWKNHGMSPMSTLOSHICXMKKEEL 1860  
Qy 1861 TSHQSOLTAFFLEALDFAQHSSENDLEEVGKTENCIIICLIVAMVVKLSEVTFPLPFKLF 1920  
Db 1861 TSHQSOLTAFFLEALDFAQHSSENDLEEVGKTENCIIICLIVAMVVKLSEVTFPLPFKLF 1920  
Qy 1921 DWAKTEDAPKDRLLTFYNLADCTIAEKLKGLFTLFAHGVKPPADTLNQVNISKTDPAFFD 1980  
Db 1921 DWAKTEDAPKDRLLTFYNLADCTIAEKLKGLFTLFAHGVKPPADTLNQVNISKTDPAFFD 1980  
Qy 1981 SENDPKCCLLOFILLNCLYKIFLFDTOHFISKERAXALMPLVDOLENRGLGEEKFOER 2040  
Db 1981 SENDPKCCLLOFILLNCLYKIFLFDTOHFISKERAXALMPLVDOLENRGLGEEKFOER 2040  
Qy 2041 VTGHLPICTAQFSVAMADSLWKPLNYQILLKTRDSSPKVRFAALITVLALAEKLENYI 2100  
Db 2041 VTGHLPICTAQFSVAMADSLWKPLNYQILLKTRDSSPKVRFAALITVLALAEKLENYI 2100  
Qy 2101 VLLPESIPFLAEIMEDECEVEHQCKTIQOELTVLGEPLQSYF 2144  
Db 2101 VLLPESIPFLAEIMEDECEVEHQCKTIQOELTVLGEPLQSYF 2144

## RESULT 2

Q7SY48 PRELIMINARY; PRT: 2159 AA.  
ID Q7SY48  
AC Q7SY48;  
DT 01-OCT-2003 (T-EMBLrel. 25, Created)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Hypothetical protein flj10359;  
GN Names:flj10359;  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
ON NCBI\_TaxID=7955;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.J., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Murzyn D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalek A., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AB; TISSUE=Whole body;  
RA Director MGC Project;  
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC055128; AAH55128.1;  
DR ZFIN; ZDB-GENE-030131-6378; flj10359.  
DR InterPro; IPR008938; ARM.  
DR InterPro; IPR000357; HEAT.  
DR Pfam; PF02985; HEAT; 2.  
KW Hypothetical protein.  
SQ SEQUENCE 2159 AA; 242048 MW; F00DBBAID597E70B CRC64;  
Query Match 53.7%; Score 5807.5; DB 2; Length 2159;  
Best Local Similarity 53.8%; Pred. No. 3.5e-262;  
Matches 1175; Conservative 391; Mismatches 554; Indels 63; Gaps 20;  
Qy 1 MTSLAQOQLRALPOSDASLSRDEVASLLFPDKEAATIDRDTAFAGCTGLEELGIDP 60  
Db 1 MTSLAHQRLALPQNDSSLLGRKEVSVLFPDKDAASMDRSTFHALGCTGLEELMGIDA 60  
Qy 61 SFEQEPALFSQLAKTLERSVOTKAVNKOLDENISLFLIHLSPYLLKPAQKCLEWLIHR 120  
Db 61 AFSEFOETLFSQSLTLERSVQSKVKNKDDKSISLFLRLSPYFLKPKALKCIEWLLHR 120  
Qy 121 FHIHLNODSLIACVLPYHETRI FVRVIOQLLKINNSKRWFWLLPVKQSGVPLAKGTLIT 180  
Db 121 FHIHLNODSLIACVLPYHETKVVRVIOQLFKIEDPTHKHWHLHGIOKPGVPLARGTLIT 180  
Qy 181 HCYKDLGFMDFCSSLVTKSVKVAEY--PGSSAQLRVLLAFYASTIVSALVAEDVSDNI 238  
Db 181 HCYKDLGFMDFVCSMVNTSVKAYSELTRDGNCPQLRVIFSFYASTIVSALDAVEKITNSI 240  
Qy 239 IAKLPYIOKGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASQIKTLTKISLIK 298  
Db 241 IAKLPFVQLGKSNLSDYTAATYMIYVQMAVKNVMEARQLVDSLVSQLSGRTPQLIR 300  
Qy 299 DGLSLIIVLLQKQSPESLGKPPHLCNVPDLITILHGISETVDVSPLLRMYLPHLVVSI 358  
Db 301 EGLSCIIILLQKQKGVIOKTYGYLCAVPTLVSTLQSIQSTVHDIQSPVLSYLLPHLHVS 360  
Qy 359 IHHVTGEETEGMDGQIYKRHLBAITKISLKNLHLLASLLFFEEYISYSSOEBMSNKV 418  
Db 361 MTQNDQEQNEGLSDS--TGLSQSFQNLQSSNLENTAAKLLLEEVVWCN--ELPSDGI 416  
Qy 419 SILNQFLPIRLLESKYPTLDVLEHLKEIADLKQELPHQFQVSLSTSGCKYQFLAD 478  
Db 417 SALNORIQTVRLEFESRYPCALDMALENHVKNVSSDNEKNLLHQFISLTSCKYQILPE 476  
Qy 479 SDTSLMLSLNHPPLAPVRILAMNHLKKIMKTSKEGVDESFIKEAVLARLGDNDIDVLSAI 538  
Db 477 SETSLMLSLNHPPLPSVRNMAVDYLKEILNSEHNSPDEAFKDALERIKDDSPVLSAL 536  
Qy 539 SAFETPKHEFSSEVTISNLNLFQRAELSKNGSEVYLKIAADILIKELISLNDQLSNQ 598  
Db 537 KALQHHMGLMDVEDTVSSLSILLHR--IKPSADWCPVLKEAVRVLDPPRIEIGNPDLKAY 594  
Qy 599 VVVCLLPFVWINDDTESAEKIAIYLSKSGICSLHPLLRGHEEALENVISTKPKGLIG 658  
Db 595 ISWELLFPVMTTRAAPCEVQLQSAITETLLISQHLPTQGWAKVLKAVLAKTSESDDLIG 654  
Qy 659 VANQKMIELADNINLGPSSMLKVVEDLISVGEESFNLKOKVTFHVLISVLVSCSSL 718  
Db 655 VANEMLTTLIKLANMDHATKENTLENVCDILSRQSSVRDRAAFVVFSSALLQSLQSM 714  
Qy 719 KET-HFPFAIRVFSLLQKKIKKLESIVTAVETIPSEWHIELMDRGPV-ELWAHYVEELN 776  
Db 715 TESQHLHTAQSVYKLELLP-----LQAYTIQEQVSDQADECLPVCVALGEFLQKIS 768  
Qy 777 STQRVAVEDSVLFSLKFKFIALKAPKSPFGDITWNPEQLKEDSRDYLHLLIGLFEMM 836  
Db 769 CGLSAEQEQLLLSLLRLLFITTLKCPDSTFKGPMWNPKEMTTTCYLRLLCLRFDVW 828

Qy	837	LANGAD-----AVHFRVLMKLFKIVHLEDVQLPFKFCVLTWYSSUSNPLNGSVKTVLOTQ	892
Db	829	ISGASQGLAPCFRSLMOPQLQVHLNPMVLPFKFLSLSWGYNLSNGDQLDCRVSAILOQTQ	888
Qy	893	ALYVCAMLSQKTOCKHOLASISSPVVTSLLINLGSPKVEVRRRAIOCLQALSG-VASP	951
Db	889	ALYVGKAFISSQPVKTLNLLASDSPPVPSLLVVCVSGVCEVRRRAIAVLCQLSLGVSP	948
Qy	952	FYLIIDHLISKAEITSDAAYVIQDIALTIFFEL---OREKKLKGHKLSETLKNLLSCV	1007
Db	949	YHPLVEKLLKSSEIADSSYLTAQLSKFEYBEAVSRKOKNKLAS-----VEQLLOCL	1001
Qy	1008	YS--CPSYIAKMLKVQGVNGEMVLSQLLPMWAEOLLEKIQEKPAPVAKDEAMVHLITLG	1065
Db	1002	QSPFCPSVTSKTLRALQDVHGPEVLSVLLPAVERLLEOCAPDSCTFLPDEALLQLLLS	1061
Qy	1066	KYNEFSVLLNEDPKSLDIFIKAVHTTKELYAGMPTIOITALEKITKPEFAISDEKVQO	1125
Db	1062	KFSEMSAPLLVKDPRCLEVFIRALHTSARPYPTIPSFQITALEQITKPFPTAIGDEKIQO	1121
Qy	1126	KLLRLMFLDLLVNCNSHCAQTVSSVFKGISVNAEOVRIELEPPDKAKPLGTVOQRRQKM	1185
Db	1122	KILSLFDLLVGNKSPACQINSVEKTIADVCELVANELIPADKQRTATVQOTRRSQM	1181
Qy	1186	QOKKSQDLF-SVOEVGSGYWQRTVLTLELQHKKLRSPQIIVPTLFNLLSRCLPBPQE	1244
Db	1182	--RKTQDTSQAVPEESVSWPRVTLLELQHKKLRQAYLPALFNLLSRCLPFAAAE	1239
Qy	1245	QGMWETKQILISCLLNIQKLSPOGKIPKQILDEKENVELIVOCIRLSSEWQTHHA	1304
Db	1240	QENIEYTKQILITCLLVNVCQKSPGEGPIKQVLEEDKFNMLVQCVRSVSEMPQTHHA	1299
Qy	1305	LLLLGVVAGTFDPKVLHNTMSIFTFMGANVRLDDTYSFQVINKTKVWIPALIQDSGD	1364
Db	1300	LLLLGALAGTFPEKVLHNTMPTTFMGANIMRLDDTYSFQVINKTKVQAVIPALIKAEHG	1359
Qy	1365	SIEVSRNVEIIVKIIISVFDALPHVPEHRRLPILVOLVDTIGAEPFLMILLILPEQVY	1424
Db	1360	SSQSEGHMETVVAQIIHVFDALPHVPEHRRLPILSQLMSTLSPSRFLVLMLLPKQHV	1419
Qy	1425	TKTVLAAAGKDAILEADTEFWFSVCCPSVQHOQOSLNMNLTLOYLKLPKEEETIPK-	1483
Db	1420	TOTSAGATGAKEAVVERDQDFILVCCPEFVKEQTLTKILIQYLMTPQPDREEAPEKK	1479
Qy	1484	---AVSFNKSEGOEAMQVFNVEHTSKOLRHFKFLSVFMSQLSSNNFLKKVY--ES	1537
Db	1480	KPRGSAVKKDETVDL--IFSVEHSGKDLRHFKFLSISFMAQLLASDGFVGKVADCED	1537
Qy	1538	GGPEILKGLBERLLETVLGYISAVAQSMERNADKLVKFWRALSKAYDILLKVNALLPT	1597
Db	1538	ITESLQALQDQLLEVLRYIQAVARVEDNADKPTAKFWRALLSKSYDTLDKVNALLPM	1597
Qy	1598	ETIPVIRGLVGNPLPSVRKALDLNNKLOQMSWKKITVTRFLKVPDLLAIVORKKX	1657
Db	1598	DTFTVMRGLVGNQLASVRKAMELNNKLOQRTKMLKEQITALLELIGTLLSIVGRSHR	1657
Qy	1658	----EGEEEOAINQOTALYTLKLLCKNFGAENPDPFVPVLTXTAKVIAPIRKEEKNVLGS	1713
Db	1658	QVTAQEEELAINQOTALYLSKLLCRNFGSDHKEEFVPLVKNKAVELVA-DKDEEKNVMS	1716
Qy	1714	ALLCIAEVTSTLEALAIPOPLSLMPSLLTTMKNTSELVSSEVILLSAALQKVETPLPH	1773
Db	1717	ALLCVAEVTSTKALAIQPLHRLMPAVLDTLKERKOLLNNEIYLLSAVTAQORASETLPH	1776
Qy	1774	FISPYLEGILSOVHLKEKITSSEMGSAQANIRLTSKKTATTLAPRVLLPAIKTKYKQI	1833
Db	1777	FISPYLLDTILQVTRTLARLLRSCQPSVRLASLSTLTKLUPPRLVPTITTKYCSM	1836
Qy	1834	EKNWKNMGPPWMSILOEHIGAMKKEELTSHSQSLTAPFFLEALDPAHQHSENDLEEVGKTE	1893
Db	1837	VDAQONRLSPLMNLILKEHISHMDKQDLNNHQSELTSFFLSALDPAHQHCCQGLKKTAEIE	1896
Qy	1894	NCIIDDVAMVVVLSEVTFRPLFPKLPDWAKTEDAPKDRLLTFYNLADCIAEKLKGLFTL	1953
Db	1897	GCVIDCLLWIMKLSEVTFRPLFPKLPDWSKIDGASKDRLLTFYRLADRIADKLGLFVL	1956
Qy	1954	FAGHLVFPFADTLQXVNIKSTDEAFPSSENDE-----KCCLLIQFIINCLYK	2001
Db	1957	FAGLVKFPFSDLLHQLNISHTDKAFPSDESDSDSDDEADNVTKSLLLYQVYLDCLJHK	2016
Qy	2002	IFLFDTHFISKERAXALMPLVDOLNRLGEGEKFORVTKHLIPCIQAQSVAMADDSL	2061
Db	2017	IFLYDTQHFISKERADALLCPLVDOLNMLGGEETYSRITTHLVPCIAQFAVAMRDSQ	2076
Qy	2062	WKPLNYQILLKTRDSSPKVRFAALITVIALAELKENYIVLLPSIPFELABELMECEV	2121
Db	2077	WKVLYQILLKTRISSPKVRFSALVMLELAGLRENYVMVLLPETIPFELABELMECEV	2136
Qy	2122	EHQOQKTIQOLETVLGEPLOSYF	2144
Db	2137	EHQVQVQIEMETILGEPLOSYF	2159
RESULT 3			
Q6P197			
ID	Q6P197	PRELIMINARY;	PRT; 1106 AA.
AC	Q6P197;		
DT	05-JUL-2004	(TREMELrel. 27, Created)	
DT	05-JUL-2004	(TREMELrel. 27, Last sequence update)	
DT	05-JUL-2004	(TREMELrel. 27, Last annotation update)	
DE	FLJ10359	protein (Fragment).	
GN	Name=FLJ10359;		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RC	TISSUE=Lymph;		
RX	MEDLINE=28289257; PubMed=12477932; DOI=10.1073/pnas.242603899;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M.J., Udwin T.B., Toshynski S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,		
RA	Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C., Butterfield Y.S.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Schen J.E.,		
RA	Krzywinski M.I., Skalska J., Smailus D.E., Schnerch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RL	and mouse cDNA sequences."		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).		
RN	[2]		
RC	SEQUENCE FROM N.A.		
RC	TISSUE=Lymph;		
RA	Strausberg R.;		
RL	Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC065205; AAH65205.1; -.		
DR	InterPro; IPR008938; ARM.		
DR	InterPro; IPR000357; HEAT.		
DR	Pfam; PF02985; HEAT; 1.		
FT	NON_TER 1		
SQ	SEQUENCE 1106 AA; 125359 MW; 09P7CE94042302C4 CRC64;		
Query Match 51.5%; Score 5562; DB 2; Length 1106;			
Best Local Similarity 99.6%; Pred. No. 4e-251;			
Matches 1102; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			

QY 1039 BOLLKIQKEPTAVLKDAMVHLTLGKYNFVSLLNEDPKSLDIFIKAVHTTKELYAG 1098  
Db 1 BOLLKIQKEPTAVLKDAMVHLTLGKYNFVSLLNEDPKSLDIFIKAVHTTKELYAG 60

QY 1099 MPTIQTITALEKIKTPFFAAISDEKVOQKLLRMLFDLLVNCNKSCHCAQTSSVFGISVNA 1158  
Db 61 MPTIQTITALEKIKTPFFAAISDEKVOQKLLRMLFDLLVNCNKSCHCAQTSSVFGISVNA 120

QY 1159 EQVRIELEPPDKAKPLGTQVQKRQKMOQKKSQDLESQVGGSYQWQVTLILELLOHKK 1218  
Db 121 EQVRIELEPPDKAKPLGTQVQKRQKMOQKKSQDLESQVGGSYQWQVTLILELLOHKK 180

QY 1219 KLRSPQILVPTLFLNLSRCLPLPBOGNMEYTTQILISCLLNTCOKLSPDGKIPKDI 1278  
Db 181 KLRSPQILVPTLFLNLSRCLPLPBOGNMEYTTQILISCLLNTCOKLSPDGKIPKDI 240

QY 1279 DEEFNFVELIVQCIIRLSEMPOTHHALLLGTGAGIPDKVLNHNMSIFTFMGANVRD 1338  
Db 241 DEEFNFVELIVQCIIRLSEMPOTHHALLLGTGAGIPDKVLNHNMSIFTFMGANVRD 300

QY 1339 DTYSFQVINKTVKMWIPALIQSDSGDSIEVSRNVEEIVVKIISVFVDALPHVPEHRLPI 1398  
Db 301 DTYSFQVINKTVKMWIPALIQSDSGDSIEVSRNVEEIVVKIISVFVDALPHVPEHRLPI 360

QY 1399 LVQLVDTLGAELFVWILLILLFEQYVTKTVLAAAYGEKDAILEADTFWFSVCCFVSQH 1458  
Db 361 LVQLVDTLGAELFVWILLILLFEQYVTKTVLAAAYGEKDAILEADTFWFSVCCFVSQH 420

QY 1459 QIOSLMNLOVLLKPEKEETIPKAVSFNKSQSEMLQVFNVEHTSKOLRHFKELSV 1518  
Db 421 QIOSLMNLOVLLKPEKEETIPKAVSFNKSQSEMLQVFNVEHTSKOLRHFKELSV 480

QY 1519 SPMQSSNNFLKVVYSGGPEILKGLERLETLVGYISAVAQSMERNADKLTVKFWR 1578  
Db 481 SPMQSSNNFLKVVYSGGPEILKGLERLETLVGYISAVAQSMERNADKLTVKFWR 540

QY 1579 ALLSKAYDLDKVNALLPTFTFPIVIRGLVGNPLPSVRRKALDNLNNKQNTISWKTTIV 1638  
Db 541 ALLSKAYDLDKVNALLPTFTFPIVIRGLVGNPLPSVRRKALDNLNNKQNTISWKTTIV 600

QY 1639 TRFLKLVDPDLALVORKKEGEBEQAINROTALYTLKCKNFGAENPDVFPVLATAVK 1698  
Db 601 TRFLKLVDPDLALVORKKEGEBEQAINROTALYTLKCKNFGAENPDVFPVLATAVK 660

QY 1699 LIAPERKEKNVLGSALECIATVSTLEALAIQPLSPMLSLTTWKNTSELVSSEVYLL 1758  
Db 661 LIAPERKEKNVLGSALECIATVSTLEALAIQPLSPMLSLTTWKNTSELVSSEVYLL 720

QY 1759 SALAALQKVVTLPHTFISPYLEGILSOVTHLEKITSEMGSASQANIRLTSKKTATTLA 1818  
Db 721 SALAALQKVVTLPHTFISPYLEGILSOVTHLEKITSEMGSASQANIRLTSKKTATTLA 780

QY 1819 PRVLLPAIKTKYKQIEKNWKNHGPMSILOEHI GYMKKEELTSHOSQLTAFPLEALDPR 1878  
Db 781 PRVLLPAIKTKYKQIEKNWKNHGPMSILOEHI GYMKKEELTSHOSQLTAFPLEALDPR 840

QY 1879 AQHSENDLEEVGKTENCIIDCLVAMVVKLSEVTFRPLFFKLFDWAKTEDAPKDRLLTFYN 1938  
Db 841 AQHSENDLEEVGKTENCIIDCLVAMVVKLSEVTFRPLFFKLFDWAKTEDAPKDRLLTFYN 900

QY 1939 LADCIAEKLGKLTPLFAGHLVKPFDATLQVNTISKTDEAPFDSSENDPEKCLLQILNLC 1998  
Db 901 LADCIAEKLGKLTPLFAGHLVKPFDATLQVNTISKTDEAPFDSSENDPEKCLLQILNLC 960

QY 1999 LYKIFLFDQTHFTISKERAXALMPLVDLENRLGGEEKQERTVKHLIPCIQAQFSVAMAD 2058  
Db 961 LYKIFLFDQTHFTISKERAXALMPLVDLENRLGGEEKQERTVKHLIPCIQAQFSVAMAD 1020

QY 2059 DSLWKPLNTQILLKTRDSDSPKVRFAALITVLAALAEKLEKXENYIVLLPESIPFLAELMEDEC 2118  
Db 1021 DSLWKPLNTQILLKTRDSDSPKVRFAALITVLAALAEKLEKXENYIVLLPESIPFLAELMEDEC 1080

QY 2119 EEVEHQCKTIQQLTETVLGEPLQSYF 2144  
Db 1081 EEVEHQCKTIQQLTETVLGEPLQSYF 1106

RESULT 4  
BP28 MACFA STANDARD; PRT; 958 AA.  
AC Q9GM44;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Protein BAP28 (OnpA-17571) (Fragment).  
GN Name=BAP28;  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Macaca.  
OX NCBI\_TaxID=9541;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Osada N., Hida M., Kuseida J., Tanuma R., Iseki K., Hirai M., Terao K.,  
RA Suzuki Y., Sugano S., Hashimoto K.;  
RT "Isolation of full-length cDNA clones from macaque brain cDNA  
libraries";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to the BAP28 family.  
CC -1- SIMILARITY: Contains 1 HEAT repeat.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB049842; BAB16728.1; ALT\_INIT.  
DR InterPro; IPR008938; ARM.  
DR InterPro; IPR000357; HEAT.  
DR PROSITE; PFS00077; HEAT\_REPEAT; FALSE\_NEG.  
FT NON\_TER 1  
FT REPEAT 920 956 HEAT.  
SQ SEQUENCE 958 AA; 108644 MW; 3DBD95C3623CFB31 CRC64;  
  
Query Match 43.8%; Score 4729; DB 1; Length 958;  
Best Local Similarity 97.5%; Pred. No. 2.5e-212;  
Matches 935; Conservative 14; Mismatches 8; Indels 2; Gaps 2;  
  
QY 1187 QKKSQDLESQVGGSYQWQVTLILELLOHKKKLRSPQILVPTLFLNLSRCLPLPQEQG 1246  
Db 1 QKKSQDLESQVGGSYQWQVTLILELLOHKKKLRSPQILVPTLFLNLSRCLPLPQEQG 60

QY 1247 NMEYTKQILISCLLNTCOKLSPDGKIPKDI LDEEFNFVELIVQCIIRLSEMPOTHHALL 1306  
Db 61 NMEYTKQILISCLLNTCOKLSPDGKIPKDI LDEEFNFVELIVQCIIRLSEMPOTHHALL 120

QY 1307 LLGTGAGIPDKVLNHNMSIFTFMGANVRDLDYTFQVINKTVKMWIPALIQSDSGDSI 1366  
Db 121 LLGTGAGIPDKVLNHNMSIFTFMGANVRDLDYTFQVINKTVKMWIPALIQSDSGDSI 180

QY 1367 EVSRNVEEIVVKIISVFVDALPHVPEHRLPIILVQLVDTLGAELFVWILLILELLOHKK 1426  
Db 181 EVSRNVEEIVVKIISVFVDALPHVPEHRLPIILVQLVDTLGAELFVWILLILELLOHKK 240

QY 1427 TVLAAAYGEKDAILEADTFWFSVCCFVSQVHQIQSLMNLQVLLKPEKEETIPKAVS 1486  
Db 241 TVLAAAYGEKDAILEADTFWFSVCCFVSQVHQIQSLMNLQVLLKPEKEETIPKAVS 300

QY 1487 FNKSESQEMQLQVFNVEHTSKOLRHFKELSVSPMSQLSSNNFLKVVYSGGPEILKGL 1546  
Db 301 FNKSESQEMQLQVFNVEHTSKOLRHFKELSVSPMSQLSSNNFLKVVYSGGPEILKGL 359

QY	1547	EBRLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDLDKVNALLPTFTFPIRVG	1606
DB	360	EBRLLETVLGYINAVQAQSMERNADKLTVKFWRALLSKAYDLDKVNALLPTFTFPIRVG	419
QY	1607	LGVNPLPSVRRKALDNLNKLQONISWKTTIVTRFLKLPDLDLAIIVORKKKEGEBEQAIN	1666
DB	420	LGVNPLPSVRRKALDNLNKLQONISWKTTIVTRFLKLPDLDLAIIVORKKKEGEBEQAIN	479
QY	1667	ROTALYTLKLLCKNFGAENPDPPVPLXTAVKLIAPERKEKNVLSALLCIAEVTSTLE	1726
DB	480	ROTALYTLKLLCKNFGAENPDPPVPLXTAVKLIAPERKEKNVLSALLCIAEVTSTILO	539
QY	1727	ALAIPOPLSMPLSLTTMTKNTSELVSSEVYLLSALAALQKVETLPHFISPYLEGILSOV	1786
DB	540	ALAVPOLPSLMPSLTTMTKNTSELVSSEVYLLSALAALQKVETLPHFISPYLEGILSOV	599
QY	1787	IHLEKITSMSGSA-SOANIRLTSKKTATTATLAPRVLLPAIKTKYKQIEKNWNHMGPPM	1845
DB	600	IHLEKITSVGSASSOANIRLTSKKTATTATLAPRVLLPAIKTKYKQIEKNWNHMGPPM	659
QY	1846	SILQEHIGVWKKEELTSHOSQLTAFPLEALDFAHQHSENDLEEVGKTENCIIIDCLVAMVV	1905
DB	660	SILQEHIGVWKKEELTSHOSQLTAFPLEALDFAHQHSENDLEEVGKTENCIIIDCLVAMVV	719
QY	1906	KLSEVTFRPLFKLFDWAKTEDAPKDRLLTFYNLADCIABKLGFTLPAHGLVKPFADT	1965
DB	720	KLSEVTFRPLFKLFDWAKTEDAPKDRLLTFYNLADCIABKLGFTLPAHGLVKPFADT	779
QY	1966	LQOVNISKTDEAPFDSNDPEKCLLLOFLNCLYKIFLFDQHFISKERAALMPLVD	2025
DB	780	LQOVNISKTDEAPFDSNDPEKCLLLOFLNCLYKIFLFDQHFISKERAALMPLVD	839
QY	2026	QLENRLGGEGKFOERYTKHLIPICIAOFSVAMADDSLWKPLNYOILLKTRDSSPKVRFAL	2085
DB	840	QLENRLGGEGKFOERYTKHLIPICIAOFSVAMADDSLWKPLNYOILLKTRDASPKVRFAL	899
QY	2086	ITVALAEALKENYIVLLPESIPFLAELMEDECEVEHQCKTIOQLETVLGSPLOSIF	2144
DB	900	ITVALAEALKENYIVLLPESIPFLAELMEDECEVEHQCKTIOQLETVLGSPLOSIF	958
RESULT 5			
Q8NTL7			
ID	Q8NTL7	PRELIMINARY; PRT; 897 AA.	
AC	Q8NTL7;		
DT	01-OCT-2002 (T-EMBLrel. 22, Created)		
DT	01-OCT-2002 (T-EMBLrel. 22, Last sequence update)		
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)		
DE	Hypothetical protein FLJ40893.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Uterus;		
RX	PubMed=14702039; DOI=10.1038/ng1285;		
RA	Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,		
RA	Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,		
RA	Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,		
RA	Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,		
RA	Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,		
RA	Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,		
RA	Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,		
RA	Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,		
RA	Tanai H., Kimoto M., Watanabe M., Hiraoka S., Chiba Y., Ishida S.,		
RA	Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,		
RA	Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,		
RA	Togiani S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,		
RA	Muashino K., Yuuki H., Oshima A., Sasaki N., Aotseuka S.,		
RA	Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,		
RA	Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,		
RA	Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,		
RA	Highigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,		
RA	Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,		
RA	Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,		
RA	Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,		
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,		
RA	Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,		
RA	Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Komatsu T.,		
RA	Togashi T., Oyama M., Hata H., Watanabe M., Takahashi Y., Nakagawa K.,		
RA	Mizushima-Sugano J., Satoh T., Shirai Y., Masuho Y., Yamashita R.,		
RA	Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Sugano S.,		
RA	Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,		
RT	"Complete sequencing and characterization of 21,243 full-length human		
RT	CDNAs."		
RL	Nat. Genet. 36:40-45 (2004).		
DR	EMBL; AK098212; BAC05261.1; --		
DR	InterPro; IPR008938; ARM.		
DR	InterPro; IPR000357; HEAT.		
PFam	PF02985; HEAT; 1.		
SQ	SEQUENCE 897 AA; 101574 MW; 5FF6A94FE8855895 CRC64;		
Query Match	41.6%; Score 4494; DB 2; Length 897;		
Best Local Similarity	99.2%; Pred. No. 2.1e-201;		
Matches 890; Conservative	2; Mismatches 5; Indels 0; Gaps 0;		
QY	1248	MEYTKQLILSCLNLT COKLSPDGGKIPKIDLBEEKFNVELIVQCIRLSEMPQTHHALLL	1307
DB	1	MEYTKQLILSCLNLT COKLSPDGGKIPKIDLBEEKFNVELIVQCIRLSEMPQTHHALLL	60
QY	1308	LGTVAGIPDKVLHNMISITFTPMGANVMRLDDTYSFQVINKTKVMVIPALIOSDSDSIE	1367
DB	61	LGTVAGIPDKVLHNMISITFTPMGANVMRLDDTYSFQVINKTKVMVIPALIOSDSDSIE	120
QY	1368	VSRNVEEIVVKIISVFDALPHVPEHRLPILVQLVDTLGAEKFLWILLILFEQVVTKT	1427
DB	121	VSRNVEEIVVKIISVFDALPHVPEHRLPILVQLVDTLGAEKFLWILLILFEQVVTKT	180
QY	1428	VLAAYGEKDAILEADTEFWSCFVSQHOI QSLMNTLOVLLKLPESKEETIPKAVSF	1487
DB	181	VLAAYGEKDAILEADTEFWSCFVSQHOI QSLMNTLOVLLKLPESKEETIPKAVSF	240
QY	1488	NKSEOSEMLOVFNVEHTSKQLRHFPLSVFMSQLISSNNPILKVVESGGPEILKGLGE	1547
DB	241	NKSEOSEMLOVFNVEHTSKQLRHFPLSVFMSQLISSNNPILKVVESGGPEILKGLGE	300
QY	1548	ERLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDLDKVNALLPTFTFPIRVGL	1607
DB	301	ERLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDLDKVNALLPTFTFPIRVGL	360
QY	1608	VGNPLPSVRRKALDNLNKLQONISWKTTIVTRFLKLPDLDLAIIVORKKKEGEBEQAIN	1667
DB	361	VGNPLPSVRRKALDNLNKLQONISWKTTIVTRFLKLPDLDLAIIVORKKKEGEBEQAIN	420
QY	1668	QTALYTLKLLCKNFGAENPDPPVPLXTAVKLIAPERKEKNVLSALLCIAEVTSTLEA	1727
DB	421	QTALYTLKLLCKNFGAENPDPPVPLXTAVKLIAPERKEKNVLSALLCIAEVTSTLEA	480
QY	1728	LAIPOLPSLMPSLTTMTKNTSELVSSEVYLLSALAALQKVETLPHFISPYLEGILSOVI	1787
DB	481	LAIPOLPSLMPSLTTMTKNTSELVSSEVYLLSALAALQKVETLPHFISPYLEGILSOVI	540
QY	1788	HLEKITSMSGSAOANIRLTSKKTATTATLAPRVLLPAIKTKYKQIEKNWNHMGPPMSI	1847
DB	541	HLEKITSMSGSAOANIRLTSKKTATTATLAPRVLLPAIKTKYKQIEKNWNHMGPPMSI	600
QY	1848	LQEHIGVWKKEELTSHOSQLTAFPLEALDFAHQHSENDLEEVGKTENCIIIDCLVAMVVKL	1907
DB	601	LQEHIGVWKKEELTSHOSQLTAFPLEALDFAHQHSENDLEEVGKTENCIIIDCLVAMVVKL	660
QY	1908	SEVTFRPLFKLFDWAKTEDAPKDRLLTFYNLADCIABKLGFTLPAHGLVKPFADTFLX	1967
DB	661	SEVTFRPLFKLFDWAKTEDAPKDRLLTFYNLADCIABKLGFTLPAHGLVKPFADTFLD	720



```

QY 1968 QVNISKTDAPFDSNDPEKCCLLQFILNCLYKIFLDTQHFISKERAXALAMPLVDQL 2027
Db 721 QVNISKTDAPFDSNDPEKCCLLQFILNCLYKIFLDTQHFISKERAGALAMPLVDQL 780
QY 2028 ENRIGGEKEQERVTHLIPCIAPQSVAMADDSLWKPLNYQILLKTRDSSPKVRFAALIT 2087
Db 781 VNRIGGEKEQERVTHLIPCIAPQSVAMADDSLWKPLNYQILLKTRDSSPKVRFAALIT 840
QY 2088 VLALAEKIKENYIVLLPESIPFLAELEMEDECEVEHOCQKTIQOELTVLGEPLQSYF 2144
Db 841 VLALAEKIKENYIVLLPESIPFLAELEMEDECEVEHOCQKTIQOELTVLGEPLQSYF 897

RESULT 6
Q7T152 PRELIMINARY; PRT; 1336 AA.
AC Q7T152;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SI:zC146F4.2.1 (Novel protein similar to human BAP28) (Fragment).
GN Name=SI:zC146F4.2;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN 1
RP SEQUENCE FROM N.A.
RA Garner P.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL732629; CAE17603.1; -
DR InterPro; IPR008938; ARM.
FT NON TER 1
SQ SEQUENCE 1336 AA; 150326 MW; AA949557F21ACBCE CRC64;

Query Match 36.4%; Score 3935.5; DB 2; Length 1336;
Best Local Similarity 58.6%; Pred. No. 4.3e-175;
Matches 794; Conservative 220; Mismatches 297; Indels 45; Gaps 13;

QY 812 WNNPEOLKEDSRDYLHLLIGLPEMMLNGAD----AVHFRVLMKLFKXVHLEDVQFLPKFC 867
Db 3 WNNPEKMETTCYCLRELLCRFLDVIIVSGASQGPLFCFRSLMQPLQVHLNEMVLFKFL 62
QY 868 SVLWYTGSSLSNPLNGSVKTVLTQALYVGCAMLSQKTCQKHQLASISPPVTSLLNL 927
Db 63 SLLWGYNSLGDQDCRVSAILOQALYVGFKAFLLSQPVKTLNLLASDSPVVPVSLVCV 122
QY 928 GSPVKEVRRAAIOCLQALSG-VASPPYLIIDHLISKAEBITSDAAYVIOQLATLPEEL-- 984
Db 123 CSGVCEVRRAAIAVLQCLSGVSPVHPLVEKLLKSSEIIADSSVLTQALSIFYEAVS 182
QY 985 --QREKKLKHQKLSLTKNLILSCVYS--CPSYIAKDLMKVLQGVNGEVMVLSQLPMASQ 1040
Db 183 RKDXNNKKLAS-----VQLQCLQSPFCPSYTKTLRLAQDVHGEFVSLVPAVER 235
QY 1041 LLEKIQKEPTAVLKDAMVHLITLKGYNFVSVLNEDPKSLDIFIKAVHTTKELVAGMP 1100
Db 236 LLEQCAPDSCFTLPDEALLQLLLSKFSEMSAPLIVKDRCLVFRALHTSARPYPTIP 295
QY 1101 TQITALEKITPFFAAISDEKVOQKLLRMLFDLLVNCNKHCAQTQSVSPFKGISVNASQ 1160
Db 296 SFQITALEKITPFFTAIGDEKIQKILSILFDLLVGNKSPACAQSGINSVFVKTIAVDCBL 355
QY 1161 VRIELEPPDKAPLGVQKRRQKMOOKSODLE-SVOEVGSGYQWRVTLIILELQHKK 1219
Db 356 VANELIPADQKRVATVQQTRSKM--RTQDTSGVAPBESVVSVPVRLIILELQHKK 413
QY 1220 LRSPQILVPTLNLNLSRCLLEPLPQEOGNMEYTKQLILSCLLNICOKLSPDGKIPKIDLD 1279
Db 414 LKRAQVLPALFNLSRCLLEPAAEQENIEYTKQLILCLLNVCKLSPEGGFISKDVLE 473

```

```

QY 1280 EEFNFVELIVQIRLSEMPQTHHALLILGTAGIFDPKVLVHNLMSITFTMGANVMRLDD 1339
Db 474 EDRFNVELVQVRVSEMPQTHHALLILGTAGIFDPKVLVHNLMSITFTMGANVMRLDD 533
QY 1340 TYSFQVINKTVKVIIPALIOSDSDSIEVSRNVEEIVKIIISVFDALPHVPEHRRPIL 1399
Db 534 TYSFQVINKTVQAVIPALIKAHGSGSSQSEGHMETVVAQIHHVFDALPHVPEHRRPIL 593
QY 1400 VQLVDLTGAEKFFWILLILFEQVTKTLLAAAYGEKDAILEADTEFWFSVCCFESVQHQ 1459
Db 594 SQLMSTLIGSRFLWVLMMLLFKHQVTSAGATGAKEAVVERDQDFWILVCCFEFVKEQ 653
QY 1460 IQSLMILQYLLKLPEKESETIPK-----AVSFNKGESQBEMLQVFNVTHTSKQLRHF 1514
Db 654 LLSLIKILQYLMTLPODREAEPEKKPRGRSAVKKDETVDL--IFSVEHSGKDLRHF 711
QY 1515 FLSVSFMSQLLSSNNFLKKVV--ESGPEILKGLBERLLETVLGYISAVASQSMERNADKL 1572
Db 712 FISISFMAQLLASDGFVGKVDCEITESTLQALQDQLLVEVLRYTQAVARCVEDNADKP 771
QY 1573 TVKFWALLSKAYDILLDKVNALLPTETFTPIVIRGLVGNPLPSVRRKALDNLNKLQONIS 1632
Db 772 TAKFWALLSKSYDILLDKVNALLPMDFTFTVMRGLMGNQLASVRRKAMELLNKLQORTK 831
QY 1633 WKTIIVTRFLKLVDPDLAIIVQRKK---EGEBEQAINRQTALYTLKLCKNFGAENPDP 1688
Db 832 WLKEQITALLELIGTLLSIVGRSHRQVTAQEEBELAINRQTALYSLKLKCRNFGSDHKEE 891
QY 1689 FVPVLTAVKLTAPERKEKNVLSGALLCIAEVTSTLEALAIQOLPSMLSLTTMKNTS 1748
Db 892 FVPVLNKAVELVA-DKDEEKNVMSGALLCVAEVTSTLKALAIQOLHRLMPAVLDTLKERK 950
QY 1749 ELVSSEVYLLSALAALQKVVETLPHFISPVLEIGILSQVTHLEKITSEMGSASQANRLTS 1808
Db 951 DLLNNEIYLLSAVLAQRASETLPHFISPYLLDTILOVTRTLILARRLTSQPOLSVRLAS 1010
QY 1809 LKKTTLATLAPRVLLPAIKTKYQIEKNWNKMHGPFMSILOEIGHMKKEBELTSHOSQLT 1868
Db 1011 LSGSTLTKLPVRVLIPTITKCYCSMVDAQNRLSPLMNILKEHISHMDKQDLNHSBELT 1070
QY 1869 APFLEALDFAQHSNDLEEVGTENCIIIDCLVAMVVKLSEVTFRPLFPKLFOWAKTEDA 1928
Db 1071 SFFLSALDFAQHCOGDLKKTABIEGCVIDCLLVIMIKLSEVTFRPLFPKLFOWSKIDGA 1130
QY 1929 PKDRLLTFYNLADCAEKLGLFTLPAHLVXKPFADTLQXVNIKSKDEAFFDSDENPEKC 1988
Db 1131 SKDRLLTFRLADRLADKLGLFVLPAQOLVXKPFADTLQXVNIKSKDEAFFDSDENPEKC 1180
QY 1989 CLLLOFILNCLYKIFLDTQHFISKERAXALAMPLVDQLENRLGGEKEQFQERVTHLIPC 2048
Db 1181 SLLQVVLDCDLHKIFLYDTQHFISKERADALLCPLVDQLENRLGGEKEQFQERVTHLIPC 1240
QY 2049 IAQFSVAMADDSLWKPLNYQILLKTRDSSPKVRFAALITVLALAEKIKENYIVLLPESIP 2108
Db 1241 IAQFAMVRDSSQWVLYNYQILLKTRHSSPKVRFSALVMLELAGLAKLRENYMVLLETIP 1300
QY 2109 FLAELMEDECEVEHOCQKTIQOELTVLGEPLQSYF 2144
Db 1301 FLAELMEDECEVEHOCQKTIQOELTVLGEPLQSYF 1336

RESULT 7
Q7T153 PRELIMINARY; PRT; 1278 AA.
AC Q7T153;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SI:zC146F4.2.2 (Novel protein similar to human BAP28) (Fragment).
GN Name=SI:zC146F4.2;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

```

OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]\_SEQUENCE FROM N.A.  
RA Garner P.;  
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL73629; CAB17602.1; -;  
FT InterPro; IPR008938; ARM.  
FT NON TER 1  
SQ SEQUENCE 1278 AA; B6C9FC81B77EE1A9 CRC64;

Query Match 34.0%; Score 3669.5; DB 2; Length 1278;  
Best Local Similarity 57.6%; Pred. No. 1e-162;  
Matches 744; Conservative 211; Mismatches 292; Indels 45; Gaps 13;

QY 812 WNPPEKEDSDRYLHLLGLFEMMLNGAD----AVHFRVLMKLFVKVHLEDFVQLFKPC 867  
DB 3 WNPPEKMETTCYLRLLCCLFDVVISGASQGPLAPCFRSLMQPLQVHLNRPWLFKFL 62

QY 868 SVLWTVGSSLNPLNSKVTVLQTOALYVGCAMLSQKTOCKHQHLASISSPVVTSLLNL 927  
DB 63 SLLWGYNSNLGDLDCRVSAILQTOALYVYKAPFLSSQPVKTLNLLASDSPVPSLLVCV 122

QY 928 GSPVKEVRRAATOCALSG-VASPEYLIIDHLISKAEEITSDAAVVIDLATLPEEL-- 984  
DB 123 CSGVCEVRRAATAVLQCLSLGVSSPHVPEVKLKSSEIIADSSVLTQALSKFYEEAVS 182

QY 985 --QREKKLSHQKLSLTKNLLSCVYS--CPSYIAKDLMKVLQGVNGEMVLSQLLPMAEQ 1040  
DB 183 RKDKNKKLAS-----VQLQCLQSPFCPSYTSKTLRALQDVHGEVPSVLLPAVER 235

QY 1041 LLEKIOKEPTAVLKDAMVHLHVLTKYKNFVSLLNEDPKSLDIFTKAVHTTKELYAGMP 1100  
DB 236 LLEQCAPDSCTFLPDEALLQLLSKFSEMSAPLLVKDPRCLEVFTRALHTSARPTPTP 295

QY 1101 TQITALEKTKTFFPAASDEKVOOKLMLPDLVNCNKSCHCAQTVSSVFKGISYNASQ 1160  
DB 296 SFQITALEKTKTFFPAIDGDKTQKILSLFDLLVGNKSPACQINSVFKTIADVCEL 355

QY 1161 VRIELEPPDKAKPLGTVQKRRQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1219  
DB 356 VANELIPADKQRTATVQOTRSEK--RKQDTSGAVPEESVVSFVRLVLELLQHKK 413

QY 1220 LRSPQILVPTFLNLSRCLLEPLPOEQGNMYTKOLISCLLNTCOKLSPDGKIPKXILD 1279  
DB 414 LKRAQVLPALFNLRLSCLPEAAAEQENIEYTKQLILICLLNVCKLSPGEGPISKDVLE 473

QY 1280 EKFNVVELIVOCIRLSEMPOTHHALLLLGTVAGIPDPKVLHNMISTFTFMGANVRLDD 1339  
DB 474 EDKFNELVVQCRVSEMPOTHHALLLLGTVAGIPDPKVLHNMISTFTFMGANVRLDD 533

QY 1340 TYSFQVINKTKVMVIPALIQSDGSDSTEVSERNVEEIVVKIISVFDALPHVPEHRLPIL 1399  
DB 534 TYSFQVINKTVQAVIPALIKAEHGGSSQSEGHMETVVAQIIHFVDALPHVPEHRLPIL 593

QY 1400 VQLVDTLGAKFVILWILLILPEQVYTKVLAAYGEKDAILEADTEFWFVSCFESVQHQ 1459  
DB 594 SQLMSTLGPFRFLVWMLLLFKQVTVTSAGATGAKEAVVERDQDFWILVCCFEFEVKEQ 653

QY 1460 IQSLMNLVLLKLPKEKETIPK-----AVSFNKSQSEEMQVFNVTHTSKQLRHFK 1514  
DB 654 LTSILKTLQVLMTLQPDREEAPEKKPRGSVAVKQDFTVSDL--IFSVEHSGKOLRHFK 711

QY 1515 FLVSFMSQLSSNFKLVV--BSGGPEILKGLERLLETVLGYISAVAQSMERNADKL 1572  
DB 712 FISISFNAQLLASDGFVKVADCEDEITESTLQALQDQLLEVLYRIQAVARCVEDNADKP 771

QY 1573 TVKFWALLSKAYDLLDKVNALLPTETPIPIVIRGLVGNPLPSVRKALDNLNKLQONIS 1632  
DB 772 TAKFWALLSKSYDTLQVALLPMDTFTITMRGLMGNQLASVRRKAMELNKLQORTK 831

QY 1633 WKKTIVTRFLKVPDLALVORKKK---EGEEEOALNROTALYTLKLLCKNFGAENPDP 1688

832 WLKEQITALLELIGTILLSIVGRSHQVTAQEEBELAINROTALYSLKLLCRNFGSDHKEE 891  
1689 FVPVLXTAVKLIAPERKEEKVLSALLCIAEVTSTLEALAIPOPLSPMLSLTTMKNTS 1748  
892 FVPVLNKAVELVA-DKDEKKNVMSALLCVAEVTSTKALAIPOPLHRLMPAVLDTLKERK 950  
1749 ELVSEVYLLSALAALOKVWVTLPHFIPSPYLEGILSQVHLEKITSFEMGASQANIRLTS 1808  
951 DLLNNEIYLLSAVTLAQASSETLPHFISPYLLDTLQVTRTLTLAARLTSCPLSVRLAS 1010  
1809 LKKTTLATLAPRVLLPAIKTKYQIEKNKNHMGFMFSILOSHIEXMKKEELTSQSOLT 1868  
1011 LSLSTLTKLPPEVLIPTITTKCYCMVDAQOQRLSPLMNLKEHISHMDKQDLNNHQSLE 1070  
1869 AFFLEALPRAQSENDELEEVGKTENCILDCIVAMVVKLSEVTFRPLPKLFDWAKTEDA 1928  
1071 SFPLSALDFRAHQCGDILKKTAEIEGCVIDCLVLMIMKLSSEVTFRPLPKLFDWAKTEDA 1130  
1929 PKDRLLTFYNLADCTAEKLGFLTFAGHLVKFPFADTLXQVNIKSTDEAFFSENDPEKC 1988  
1131 SKDRLLTFYRLADRIADKLKGLFVLFAQOLVKFPFSDLLHQLNTSHTGKS-----S 1180  
1989 CILLOFIILNCLYKIFLFTQHFISKERAXALMPLVDLENLGGEEKFQERTKHLIPC 2048  
1181 SLLQVLDLCLHKIFLYDTQHFELSKERADALLCPLVDQLENLGGEEKFQERTKHLIPC 1240  
2049 IAQFSVAMADDSIMKPLNYQIILLKTRDSSPKV 2080  
1241 IAQFAMVRDDSQWVLYNYQIILLKTRHSSPKV 1272

RESULT 8  
Q8BLJ4 PRELIMINARY; PRT; 733 AA.  
AC Q8BLJ4;  
DT 01-WAR-2003 (TrEMBLrel. 23, Created)  
DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length  
DE enriched library, clone:BI3001612 product:similar to PROTEIN BAP28  
DE (Fragment).  
DE Name=BI3001612Rik;  
DE Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RL Meth., Enzymol. 303:19-44 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA RIKEN FANTOM Consortium;  
RL "Functional annotation of a full-length mouse cDNA collection.";  
RN Nature 409:685-690 (2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;  
RA The FANTOM Consortium,  
RL "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs.";  
RN Nature 420:563-573 (2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,

RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RN Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuana T., Tashiro H., Itoh M.,  
 RA Sui N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuiura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RN Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayaehida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tomaru A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK049699; BAC32161.1; -  
 DR MGD; MGI:2442524; B13001612Rik.  
 DR InterPro; IPR008938; ARM.  
 DR NeoPro; IPR000357; ARM.  
 DR Pfam; PF02985; HEAT; 2.  
 DR NON TER 733 733  
 SQ SEQUENCE 733 AA; 81938 MW; 513394872B6250EC CRC64;  
 Query Match 28.7%; Score 3099; DB 2; Length 733;  
 Best Local Similarity 81.3%; Pred. No. 2.1e-136;  
 Matches 596; Conservative 73; Mismatches 64; Indels 0; Gaps 0;  
 QY 1 MTSLAQQLRLALPQSDASLLSRDEVASLLFDPKAATIDRDTPAFAGCTGLEELGIDP 60  
 DB 1 MTSLAQQLRLALPQSDASLLSRDEVASLLFDPKAATIDRDTPAFAGCTGLEELGIDP 60  
 QY 61 SFEQPEAPLPSQAKTLERSVQTKAVNKQDENISLFIHLSPYFLKPAQKCLEWLIHR 120  
 DB 61 AFEQPEAPLPSQAKTLERSVQTKAVNKQDENISLFIHLSPYFLKPAQKCLEWLIHR 120  
 QY 121 FHILHNQDSLACVLPIYHETRIFFVRIQLLKNNKSHRWFWLLPVKQSGVPLAKGTLLT 180  
 DB 121 FHILHNQDSLACVLPIYHETRIFFVRIQLLKNNKSHRWFWLLPVKQSGVPLAKGTLLT 180  
 QY 181 HCYKDLGFMDFICSLVTKSVKVPFAEYPGSSAQLRVLLAFVASTIVSALVAEDVSNITA 240  
 DB 181 HCYKDLGFMDFICSLVTKSVKVPFAEYPGSSAQLRVLLAFVASTIVSALVAEDVSNITA 240  
 QY 241 KLFPYIQKGLKSLPYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKDG 300  
 DB 241 KLFPYIQKGLKSLPYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKDG 300  
 QY 301 LSLVILLQRPESLGKPPFPHLCNVPDLITILHGISTDYVSPLLRMLPHLVSIITH 360  
 DB 301 LGLCLILLQRPENLGERPFLHLCGVPDLILGHGISESDYVSPLLRMLPHLVASVQ 360  
 QY 361 HVTGETECMDGQIQYKRLHLEILTKISLKNNDLHLLALLFEEYISYSSQEEEDMSKVSL 420  
 DB 361 HIAGEBEAGIDGQIQYKRLHLEILTKISLKNNDLHLLALLFEEYISYSSQEEEDMSKVSL 420

421 LNEQFLPLRLLESKYPRTLDVWLEHLKEIADLKKQELPHQFVSLSTSGKYQFLADSD 480  
 DB 421 LNEQFLPLRLLESKYPRTLDVWLEHLKEIADLKKQELPHQFVSLSTSGKYQFLADSD 480  
 QY 481 TSLMLSLNHPPLAPVRILAMNHLKKIMKTSKEGVDESIFKEAVLARLGGDDNIDVLSAISA 540  
 DB 481 TSLMLSLNHPPLAPVRILAMNHLKKIMKTSKEGVDESIFKEAVLARLGGDDNIDVLSAISA 540  
 QY 541 FEIFKEHFSSEVITISNLLNLFQRAELSKNGEWEVLKIAADILITKEILSENDOLSNOVV 600  
 DB 541 FEIFQHQHFGVEETISNLLNLFQRAELSKNGEWEVLKIAADILITKEILSENDOLSNOVV 600  
 QY 601 VCLLPFWVINDDTESAEMKIAIYLSKSGICSLHPLLRGWEEALENVIKSTKPKGLIGVA 660  
 DB 601 VQLLPFWVITNSDIESPDMKIAIHLKSGICSLHPLLRGWEEALENVIKSTKPKGLIGVA 660  
 QY 661 NQKMIELLADNINLNGPSSMLKQVEDLISVGEESFNLKQVTFHVLVSVSCCSLKE 720  
 DB 661 NQKMWQLLGSNLGSRSTVLKLVEDLVACAGEKESYSKQKVAHFVTVSVLISCCSFQE 720  
 QY 721 THPPFAIRVFSLL 733  
 DB 721 TCPPFAIRVFSLL 733  
 ID Q7PW6 PRELIMINARY; PRT; 2098 AA.  
 AC Q7PW6;  
 DT 01-MAR-2004 (TREMBLrel. 26, Created)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE ENSANGP0000017181 (Pragment).  
 GN Name-ENSANG0000014692;  
 OS Anopheles gambiae str. PEST.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.  
 OX NCBI\_TaxID=180454;  
 RN [1]\_TaxID=180454;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AAAB01008984; EAA14843.1; -  
 DR InterPro; IPR001917; Aminotrans\_II.  
 DR NeoPro; IPR008938; ARM.  
 DR InterPro; IPR000357; HEAT.  
 DR Pfam; PF02985; HEAT; 3.  
 DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; UNKNOWN\_1.  
 FT NON TER 1  
 SQ SEQUENCE 2098 AA; 235889 MW; F4ED8C24CEAD0070 CRC64;  
 Query Match 19.4%; Score 2094.5; DB 2; Length 2098;  
 Best Local Similarity 28.9%; Pred. No. 6.8e-89;  
 Matches 667; Conservative 401; Mismatches 859; Indels 377; Gaps 69;  
 QY 2 TSLAQLQRLALPQSDASLLSRDEVASLLFDPKAATIDRDTPAFAGCTGLEELGIDPS 61  
 DB 11 TGLAQLKRLAAPQTSAPVDAR-RTASILFDKAEAAKDRVYIDIGVSGLEELTOMHGA 69  
 QY 62 FPEQPEAPLPSQAKTLERSVQTKAVNKQDENISLFIHLSPYFLKPAQKCLEWLIHR 121  
 DB 70 FAQFEDTLFDKAMDLOQRSVENREYNEQDGNIRRFVHLSFYFLMQLPAKCLEWLIHR 129  
 QY 122 HHLHNQDSLACVLPIYHETRIFFVRIQLLKNNKSHRWFWLLPVKQSGVPLAKGTLLT 181  
 DB 130 EHLKYNRELFWLILPFIHETRIFFVRCVQTMQIEFNDRILAFVAVKSGVPPAKRTIHD 189  
 QY 162 CYKDLGFMDFICSLVTKSVKVPFAEYPGSSAQLRVLLAFVASTIVSALVAEDVSNIIAK 241

Db 190 CVSHFGFLQXQSFITGAVE---ELAGRANALQATFAPYCTTALGMLHSAETVSNHNTA 246  
Qy 242 LFPYIQKGLSLPYRAATYMIICQISVQVMENTFVNSLASQIIKTITKIPSLIKDGL 301  
Db 247 VHTVKGGLASRAIDFAAGSPWVQGLVWKASLAQLTVDYLARRVI--AVQLPALTAET 304  
Qy 302 SCLIVLQROQ-----KPESIG-----KKPPPHCNVPPDLITILHGISYDVSPILRY 349  
Db 305 MLLVLIFOTQHERLATLSRELMEIRRCCKWLAFTLCTVK-----ADGVDVLVLYRK 355  
Qy 350 MFLPVVSIHHVTGEETGMDGOYKRLHAILTKISIKNNLDHLLASLLFEEY----- 404  
Db 356 LLEKCLNEICC-----TKGA-LKLYSGFCEOLMEIQUTEVEAEVIVQCVLDSTFKDV 408  
Qy 405 -----ISYSSQEMD-SNKVSLMNEQFLPLIRLLESKYPRTLDVVLEHLK- 449  
Db 409 PEKKAANSDDTIELDSEEDDFVSRDOHVQWYSEVYKSPEROYPGAFAKVVVRIMKG 468  
Qy 450 -BIADLKQELPHQFVSLSTSGGKYQFLADSTSLMNLNHPPLAVRILAMNHL---KKI 505  
Db 469 QOQYSPKKNALRNVLGFLQAS-----YDENETNVFESLFHYDADRRAHVOYLVENLRA 524  
Qy 506 MKTSEKGVDESGFIEKAVLARLGDNDIDV--LSAISAFIEIPEHFSSEVTISNLLNFQ 562  
Db 525 MKLKSTGQVD--LLRDSVRERLADSCVEIBELKIDPOELIVVGTDEL-IGRLSQLVL 581  
Qy 563 RAELSKNGEWEYELKIAADILIKEEILSENDSQVNVVCLLPFVV-INND----- 612  
Db 582 KCAANQT-RWAKTPRVIELLTDVVYGRGN--VNOIITIALYPLFPIGNDPVGORNKA 638  
Qy 613 -----DTESAMKIAIYLSKGICSLHPLLRGWEALENV--TKSTKP 653  
Db 639 LLSTSPAKOYLVARFYGNEENCLEMLSTLL-BAGQCT-NPV-----QVCFNILLSSALP 691  
Qy 654 GK-LIGVANQKM-----TELLADNINLGDPPSSMLKVEDLISVGE 693  
Db 692 TPCVTGQADRVLDYAVRQLQNHFRFYAPTSIDQCLQDN-----QLPQDLITVPTR 741  
Qy 694 ESNLAKQVTFHVLVS-LVSCSSLIKETHFFPAIRVFSLLQKIKKLESVITAVEIPSE 752  
Db 742 ---YILERVQFEPIGANFRQCVSLK-----LRLAIFAVILLEOYCTIDPQHAV----- 788  
Qy 753 WHIELMDRIGIPVELMAHYVEELNSTORVAVESVFLVPSLKKFIYALKAPKSPFKGDW 812  
Db 789 -----RSIFGELLKYSLSLYPO---LVDRIEFL-----SHF 817  
Qy 813 WNEQKEDSRDYLHLLIGLFEMMLNGADAVHFRVLMKLFIKVHLEDVFLQFKFCSVLT 872  
Db 818 YTAHLIEPGSKD-----DADNFLINAKQIR-----CORL-- 847  
Qy 873 YGSSLSNPLNCSVKTVLQALYVGCAMLSSQKTQCKHQLASTSSPVVTSLLINLGPVK 932  
Db 848 -----LMTLEQVSQES---GC-----TVTESALINVLISLTSGSA 881  
Qy 933 EVRAATQCLQAL-----SGVASPFYLIIDHISKAETSAAVYQDLA-TLFEELQ 985  
Db 882 VRECTMRTLIQVLGRKNGLSKAFNGFVKRMVYKREELTMDG-----EQALWVFAIFS 937  
Qy 986 REKKLSHQKLSLTKNLLSCVY--SCPSYIAKDLMKVLOVNGEMVLSQLLPWAQOLLE 1043  
Db 938 LEEBEGDGHLPQPMESFVARIWDPSTPTIYVCGIVDLLKLLDDDPHVLSCASKGVSII 997  
Qy 1044 KIQKEPTAVLKD--EAMVLHTLKGNEFSVSLNEDPKSLDIPKAVHTTKELYAGMP- 1100  
Db 998 RVNAAASQPVDAVESRVQVLLTRFNLDIVAHLPGSAICQRLVTLAIRCDKPLC--LPS 1055  
Qy 1101 -----TIQITALEKITPPFAAISDEKVOQKLRMLFDLIVNCKNSHCAOTVSVFKGIS 1155  
Db 1056 ERKLCTPAIAMEALAGEVFETL-PKAYAEVLQCVVAAATLSHPETNAATGKLFKSSA 1114  
Qy 1156 VNAQVRIEPPPKAPLGTVOQKRRQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1215  
Db 1115 IDAEVVVDMLCAMYDGRADDATGQKPGKGRKSIAPSERVLGSTWKGCVTLLEHLQ 1174

RESULT 10  
BP28\_DROME  
ID\_BP28\_DROME

STANDARD;

PRT; 2096 AA.







QY 302 SCLIVLQKQPSGLKPPHLCNVPLDITILHGISSETVDVSPLLRMLPHLVVSI-- 359  
Db 298 LLLVCIIYQKQ-----QAALPHF--KPE--TILNLVGVKMWLISTSLAKGNIAIQSICM 347  
QY 360 ----HHVTGETEGMDGOIVKRHLEAILTKISLKNLHDHLLASLLPREYIS----- 406  
Db 348 PLMTGAVAAIRDDDDASSNSCKLFDNLSEVPMKPTAQQLNCLFDTYVETAIADPEPM 407  
QY 407 -----YSSQEMDSNKVSLNEOPLPIRILLESKYPTLTDVLEELKEIADL 454  
Db 408 ETNSNEDDDTIVIDSDEIETKTT--FQAWSTVLEKLERRYPPEAFDLSVKEALRSKST 466  
QY 455 KQQLFHFQV---SLSTSGKYQFLADSDTSLMLSLNHPLAPILAMNH-LKKIMKTSK 510  
Db 467 SNKQKALKAALGFRLNTTDEKAKHAYE-----KLYHADWRLSAVQKLLQNLNLTCK 519  
QY 511 EGVDESPIKEAVLARIGDDNDIDVLSAIS-AFEIFKEHFSSEVTSINLNLFORAELSKN 569  
Db 520 RERSVKLLOECLPDRINDSGAVVSTILLSUPTTELAEMGLPPLAQTLCHLLYRAQSEKD 579  
QY 570 GEWYEVUKIAADILIKEIISENDQLSNVQVVCILLPFVWINNDDTESAEMKIAIYLSKSG 629  
Db 580 EEWQPVVPLAVRHLTSALVSGSYD--TNLVLLALMPLLPFCEALAEHOKALRILLG-SD 636  
QY 630 ICSLHPLLRGWEALENVIKTPCKLIGVANQKMIELLA-DNINLGDPDS-SMLKWVEDL 687  
Db 637 FVSKVPFLA--ELKVSNNKFDNF-----VGEHRQHFLDIIASSNQELSSQERALLQSVED- 689  
QY 688 ISVGEESFNLKOKVTHVILSVLSCSSIKETHPPFAIRVFSLLQKTKKLESVITAV 747  
Db 690 --HGGEIYQKASQLT-HILLILITAYAKRELQPSHLMLEKLGYSRRLOQ--FRVNGS 744  
QY 748 E-----IPSEWHIELMDRGIPVELMAHYVEELNSTQRAVEDSVFLVSLKFKPI---Y 798  
Db 745 QNTPNCAPLQLYVDFLLTL-VKNTKWT---ALASTPNQMTDELRLCLRLLEIICAQVF 799  
QY 799 ALKAPKSPKCDIWNNEQ-----LKEDSRDYHLHLLIGLFEMMLNGADAVHFRVLMKLF 853  
Db 800 SEKADQ--PERQWTRALQOSLOLILPEAQDRDLVLSN----- 835  
QY 854 KVHLEDVOLFCKCSVLWTYSSLSNPLNCSVKTVLQTOALYVGCAMLSQKTOCKHQLA 913  
Db 836 -----FYVFERLPELWPRDSYA-----VFLQGFILAEVLSNPKSQIDCGLV 879  
QY 914 STSPVVTSLINLNGSPVKEVRRAAIOCLQALSG--VASPFYLIIDHILISKABEITSDAA 971  
Db 880 H-----VLRVANAGSPQLTRVQAINTLQIISNRKLVSHVEQLVRSLLQKSELSNDHE 934  
QY 972 YVIQDLATLPEELOREKKLKHOKLSETLKNLLSCVVCPS-----YIAKOLMKVLQVNGE 1028  
Db 935 ---QYALILYILEPEKATKERLVLSKLKRSVLALASDPKQSPICITASLLAALKHVND 991  
QY 1029 MVLSQLPMAEQLEKIQEPTAVLXDEAMVHLTLGKYNFSPVSLN--EDPKSIDIFI 1086  
Db 992 NFLNELPLGLDLSKTI-----TAGEDNQNIKHL-PWHEIYKYSVIERFEGVVALNVLL 1045  
QY 1087 K-----AVHTTKELYAGMPTIQITALEKITKPFPAIAISDEKVOOK-----LLR 1129  
Db 1046 RKDLAWKLPEDSPAQYDVTYVQLSQKLOPLPCVLLNSITPETF-----EQMHAGKIALIK 1100  
QY 1130 MLFDLIVNKNSHCAQTVSVFKGI SVNAEQVRIELEPPDKAKPLGTVOQKRQRKMOQK 1189  
Db 1101 LIVESATNSDN-----DSIFLASHRLKRCRLDQCP--LVPILLEMANTKVEKKQPVK 1151  
QY 1190 SQDLESVQ-EVGGSYWQVRTLLELLOHKKLRSPQILVPTLFNLLSRCLLEPLPOQGNM 1248  
Db 1152 RRSVQATQDLTSPYKQWQMTLELLEHKKQLIGAELLIPPLPELLQACL--TWBESHA 1209  
QY 1249 EYTKQILSLCLNICOKLSPDGKI PKDIIIDEKFNVELIVQICIRISEMPTQHHALLLL 1308  
Db 1210 EYPKQILLSLLHCCQTAQASAGVLVK-AMPSSFRIELVQSLRNTRNPQTQOHALFL 1268  
QY 1309 GTVAGIFDPKVLNINMISITFFMGANVMRLDDTYSFQVINKTVKQVIPALIQSDSGDSIEV 1368

## RESULT 12

Q8CCT5 PRELIMINARY; PRT; 408 AA.  
ID Q8CCT5; AC Q8CCT5; DT 01-MAR-2003 (TEMBLrel. 23, Created)  
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)  
DE Mus musculus adult male olfactory brain cDNA, RIKEN full-length  
DE enriched library, clone:6430400D06 product:hypothetical ARM repeat  
DE structure containing protein, full insert sequence.  
GN Name=B13001612Rik; Synonyms=BC019693;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1] SEQUENCE FROM N.A.  
 RP STRAIN=C57BL/6J; TISSUE=Olfactory brain;  
 RC MEDLINE=9927953; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RX Carninci P., Hayashizaki Y.;  
 RA "High-efficiency full-length cDNA cloning."  
 RT Meth. Enzymol. 303:19-44(1999).  
 RL [2] SEQUENCE FROM N.A.  
 RN STRAIN=C57BL/6J; TISSUE=Olfactory brain;  
 RP MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RC RIKEN FANTOM Consortium;  
 RX "Functional annotation of a full-length mouse cDNA collection."  
 RT Nature 409:685-690(2001).  
 RL [3] SEQUENCE FROM N.A.  
 RN STRAIN=C57BL/6J; TISSUE=Olfactory brain;  
 RP The FANTOM Consortium;  
 RC "Analysis of the mouse transcriptome based on functional annotation of  
 RA 60,770 full-length cDNAs."  
 RT Nature 420:563-573(2002).  
 RL [4] SEQUENCE FROM N.A.  
 RN STRAIN=C57BL/6J; TISSUE=Olfactory brain;  
 RP MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RX Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RA "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes."  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5] SEQUENCE FROM N.A.  
 RN STRAIN=C57BL/6J; TISSUE=Olfactory brain;  
 RP MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RX Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6] SEQUENCE FROM N.A.  
 RN STRAIN=C57BL/6J; TISSUE=Olfactory brain;  
 RP Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK032134; BAC27721.1; -  
 DR MGD; MGI:2442524; BL3001612Rik.  
 DR InterPro; IPR008938; ARM.  
 KW Hypothetical protein.  
 SQ SEQUENCE 408 AA; 4668 MW; 42AB9EB13CA3FE67 CRC64;  
 Query Match 17.8%; Score 1926; DB 2; Length 408;  
 Best Local Similarity 91.9%; Pred. No. 5.2e-82;  
 Matches 375; Conservative 17; Mismatches 16; Indels 0; Gaps 0;  
 QY 1737 MPSLLTMMKNTSELVSVLSALAAQKVETLPHFISPYLEGILLQVHLEKITSEM 1796

Db 1 MPSLLTMMKNTSELVSVLSALAAQKVETLPHFISPYLEGILLQVHLEKITREM 60  
 QY 1797 GSASQANIRLTSLKKTTLATTLAPRVLLPAIKTKYKQIEKNWKNHMGPPMSILQEHGXMK 1856  
 Db 61 GSASQANIRLTALKKTTLATLSPRVLLPAISKTQKQIKNWNHMGPPMSILQEHGVYMK 120  
 QY 1857 KEELTSHOSQLTAPFLEALDFAQHSNDLBEVGTENCICIDCLVAMVVKLSEVTFRPLF 1916  
 Db 121 KEELTSHOSQLTTPFLEALDFAQHSNDLBEVGTENCICIDCLVAMVVKLSEVTFRPLF 180  
 QY 1917 FKLPDWAKTDPKDRLLTFYNLADICIAEKLGKGLFTLPAHLVKPFADTLXOVNISKTDE 1976  
 Db 181 FKLPDWAKTDPKDRLLTFYNLADICIAEKLGKGLFTLPAHLVKPFADTLXOVNISKTDE 240  
 QY 1977 AFFDSENDPEKCCLLQLFILNCLYKIFLFDQHFISKERAXALMPLVDLENRLGGBEK 2036  
 Db 241 AFFDSENDPEKCCLLQLFILNCLYKIFLFDQHFISKERAXALMPLVDLENRLGGBEK 300  
 QY 2037 FOERTVTKHLIPCIQAFSVAMADDSLWKNLYOILLKTRDSSPKVFAALITVLAELK 2096  
 Db 301 FOERTVTKYLPVCIQAFSVAMADDSLWKNLYOILLKTRDSSPKVFAALITVLAELK 360  
 QY 2097 ENYIVLLPESIPFLAELMEDECEEVEHOCQKTIQOETVLGEPLOSIF 2144  
 Db 361 ENYIVLLPESIPFLAELMEDECEEVEHOCQKTIQOETVLGEPLOSIF 408

RESULT 13  
 Q96ES5 PRELIMINARY; PRT; 349 AA.  
 AC Q96ES5;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE FLJ10359 protein.  
 GN Name=FLJ10359;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1] SEQUENCE FROM N.A.  
 RP TISSUE=Ovary;  
 RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2] SEQUENCE FROM N.A.  
 RP TISSUE=Ovary;  
 RC Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC011983; AAH11983.1; -  
 DR InterPro; IPR008938; ARM.  
 DR InterPro; IPR000357; HEAT.

```
DR Pfam; PF02985; HEAT; 1.
SQ SEQUENCE 349 AA; 39921 MW; 3A359597FF7F7079BB CRC64;
Query Match 16.5%; Score 1779; DB 2; Length 349;
Best Local Similarity 99.1%; Pred. No. 3e-75;
Matches 346; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1796 MGSASQANIRLTSLKKTATLAPRVLLPAIKTKYQIERNKNHMGPFMSIIQEHIGXM 1855
Db 1 MGSASQANIRLTSLKKTATLAPRVLLPAIKTKYQIERNKNHMGPFMSIIQEHIGXM 60
QY 1856 KKEELTSHQSLTAFFLEALDFAHQSENDELEEVGKTENCIIICLVAMVVKLSVTPRPL 1915
Db 61 KKEELTSHQSLTAFFLEALDFAHQSENDELEEVGKTENCIIICLVAMVVKLSVTPRPL 120
QY 1916 FFKLFDWAKTEDAPKORLLTFYNLADCIAGKGLFTLFAGHLVKPFPADTLXQVNISKTD 1975
Db 121 FFKLFDWAKTEDAPKORLLTFYNLADCIAGKGLFTLFAGHLVKPFPADTLXQVNISKTD 180
QY 1976 EAPFDSNDPEKCCLLLOFTINCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGEE 2035
Db 181 EAPFDSNDPEKCCLLLOFTINCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGEE 240
QY 2036 KFORVTKHLIPCIQAQSVAMADSLWKPLNYQILLKTRDSSPKVRFPAALITVLAAEKL 2095
Db 241 KFORVTKHLIPCIQAQSVAMADSLWKPLNYQILLKTRDSSPKVRFPAALITVLAAEKL 300
QY 2096 KENYIVLLPESIPFLAELMEDECEVEHOCQKTIQOELTVLGEPLOSIF 2144
Db 301 KENYIVLLPESIPFLAELMEDECEVEHOCQKTIQOELTVLGEPLOSIF 349
RESULT 14
Q8VCCK1 PRELIMINARY; PRT; 349 AA.
AC Q8VCCK1;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE B13001612Rik protein.
GN Name=B13001612Rik;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Tothiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
RA Director MGC Project;
```

```
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019693; AAH19693.1; -.
DR MGD; MGI:2442524; B13001612Rik.
SQ SEQUENCE 349 AA; 40166 MW; 9763D0331AD0F515 CRC64;
Query Match 15.6%; Score 1682; DB 2; Length 349;
Best Local Similarity 92.8%; Pred. No. 1e-70;
Matches 324; Conservative 14; Mismatches 11; Indels 0; Gaps 0;
QY 1796 MGSASQANIRLTSLKKTATLAPRVLLPAIKTKYQIERNKNHMGPFMSIIQEHIGXM 1855
Db 1 MGSASQANIRLTSLKKTATLAPRVLLPAIKTKYQIERNKNHMGPFMSIIQEHIGXM 60
QY 1856 KKEELTSHQSLTAFFLEALDFAHQSENDELEEVGKTENCIIICLVAMVVKLSVTPRPL 1915
Db 61 KKEELTSHQSLTAFFLEALDFAHQSENDELEEVGKTENCIIICLVAMVVKLSVTPRPL 120
QY 1916 FFKLFDWAKTEDAPKORLLTFYNLADCIAGKGLFTLFAGHLVKPFPADTLXQVNISKTD 1975
Db 121 FFKLFDWAKTEDAPKORLLTFYNLADCIAGKGLFTLFAGHLVKPFPADTLXQVNISKTD 180
QY 1976 EAPFDSNDPEKCCLLLOFTINCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGEE 2035
Db 181 EAPFDSNDPEKCCLLLOFTINCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGEE 240
QY 2036 KFORVTKHLIPCIQAQSVAMADSLWKPLNYQILLKTRDSSPKVRFPAALITVLAAEKL 2095
Db 241 KFORVTKHLIPCIQAQSVAMADSLWKPLNYQILLKTRDSSPKVRFPAALITVLAAEKL 300
QY 2096 KENYIVLLPESIPFLAELMEDECEVEHOCQKTIQOELTVLGEPLOSIF 2144
Db 301 KENYIVLLPESIPFLAELMEDECEVEHOCQKTIQOELTVLGEPLOSIF 349
RESULT 15
Q8T9E7 PRELIMINARY; PRT; 1690 AA.
AC Q8T9E7;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE SD03723P.
GN ORFNames=CGI0805;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Friese E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunco J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY069800; AAL39945.1; -.
DR FlyBase; FBgn0031864; CGI0805.
DR InterPro; IPR008938; ARM.
SQ SEQUENCE 1690 AA; 191874 MW; B7CE254A4FBADF09 CRC64;
Query Match 13.5%; Score 1462.5; DB 2; Length 1690;
Best Local Similarity 27.1%; Pred. No. 1.6e-59;
Matches 496; Conservative 342; Mismatches 755; Indels 237; Gaps 59;
QY 402 EYISYVSQSEMDSNKVSLLNEQFLIRLLSKYPRTLDDVLEHLKBIADKKQELFH 461
Db 9 DDTIVIDSDEIETKTT-FOAWYSTYLEKERRYPEAFDLSVKEALRSKSTSNRQKAL 67
QY 462 QFV---SLSTSGCKYQFLADSDTSLMLSLNHPPLAPVRILAMNH-LKKIMKTSKEGVDSF 517
Db 68 KLALGFRLLTDEKAKHAYE-----KLYHSADWRMSAVOKLLQNLNVTKKRSRVLK 120
```

